

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2004, 19:26:18 ; Search time 4456 Seconds
(without alignments)
11147.019 Million cell updates/sec

Title: US-10-644-659a-1

Perfect score: 1146

Sequence: 1 atggctccggggaagga.....tgattacgctactcaagtga 1146

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sv.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sta.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.man.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1146	100.0	1146	9	AF503617 Homo sapi
2	1144.4	99.9	1322	6	AX747674 Sequence
3	1144.4	99.9	1322	9	AK092694 Homo sapi
4	1142.8	99.7	2781	9	AL833422 Homo sapi
5	1141.2	99.6	2832	9	AL832152 Homo sapi
6	791.2	69.0	1128	10	AF504061 Mus muscu
7	784.8	68.5	1285	10	AF336113 Rattus no
C 8	667.4	58.2	152130	9	AP000856 Homo sapi
C 9	667.4	58.2	153147	2	AF303733 Homo sapi
C 10	667.4	58.2	186751	9	AC023344 Homo sapi
C 11	419.8	36.6	164806	10	AC138715 Mus muscu
C 12	419.8	36.6	174846	10	AC129212 Mus muscu
C 13	419.8	36.6	186137	10	AC099736 Mus muscu
C 14	373	32.5	73341	2	AC136853 Rattus no
C 15	373	32.5	220702	2	AC114181 Rattus no
C 16	227.2	19.8	145746	5	AL953858 Zebrafish
C 17	213.2	18.6	173312	2	EX324201 Danio rer
C 18	213.2	18.6	240809	2	EX571763 Danio rer
C 19	205	17.9	261	6	AX884478 Sequence
C 20	205	17.9	261	6	BD024088 Sequence
C 21	205	17.9	265	6	BD075996 5' EST of
C 22	72	6.3	7218	6	166494 Sequence 14
C 23	56.8	5.0	10732	6	E32986 Gene encodl
C 24	56.4	4.9	1943	3	AY060993 Drosophil
C 25	55	4.8	66497	2	AC015388 Drosophil
C 26	55	4.8	171451	3	AC010561 Drosophil
C 27	55	4.8	304419	3	AE003476 Drosophil
C 28	53.4	4.7	863	3	AY113283 Drosophil
C 29	52.4	4.6	22747	3	Z81532 Caenorhabdi
C 30	52.2	4.6	2000	6	AX655393 Sequence
C 31	51.8	4.5	39482	3	Z35663 Caenorhabdi
C 32	49.8	4.3	24604	3	AL009194 Drosophil
C 33	49.8	4.3	67513	2	AC017770 Drosophil
C 34	49.8	4.3	177724	3	AC104146 Drosophil
C 35	49.8	4.3	193714	3	AC104149 Drosophil
C 36	49.8	4.3	232830	3	AE003423 Drosophil
C 37	49.6	4.3	125020	9	AF429315 Homo sapi
C 38	46.6	4.1	256744	2	AC132718 Rattus no
C 39	45.6	4.0	125020	9	AF429315 Homo sapi
C 40	45.4	4.0	195539	2	AC107665 Mus muscu
C 41	44.8	3.9	173599	2	AC141388 Rattus no
C 42	43.8	3.8	231676	2	AC134063 Rattus no
C 43	43.4	3.8	121402	10	AL929064 Mouse DNA
C 44	43.4	3.8	159560	10	AL929036 Mouse DNA
C 45	43.2	3.8	195729	2	AC132940 Homo sapi

ALIGNMENTS

RESULT 1	AF503617	1146 bp	mRNA	linear	PRI 01-JUL-2002
LOCUS	AF503617				
DEFINITION	Homo sapiens STARS mRNA, complete cds.				
ACCESSION	AF503617				
VERSION	AF503617.1	GI:20530823			
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 1146)				
	Arai, A., Spencer, J.A. and Olson, E.N.				
	STARS, a Striated Muscle Activator of Rho Signaling and Serum				
	Response Factor-dependent Transcription				

Db 57 ATGGCTCCGGCGGAAAGGAGTGGGAGGGCCCAAGCAAGAGCGCCCTCCGGAAGATA 116
QY 61 CGCAGAGCCACCTGTGTATCAGCTTGGCCCGAGGTTGGCAGAGTGGCGCAATCAGAAC 120
Db 117 CGCAGAGCCACCTGTGTATCAGCTTGGCCCGAGGTTGGCAGAGTGGCGCAATCAGAAC 176
QY 121 AGCATCAGGAGCGCCAGGAGCTCAGAGCTGGCTGGCGGAGGGACCCAGAGACTCACT 180
Db 177 AGCATCAGGAGCGCCAGGAGCTCAGAGCTGGCTGGCGGAGGGACCCAGAGACTCACT 236
QY 181 CAAAGTCTTAAACCAATCAGCCCTTACTTACACAGAAAGCTCAGAGTGGCCCAAG 240
Db 237 CAAAGTCTTAAACCAATCAGCCCTTACTTACACAGAAAGCTCAGAGTGGCCCAAG 296
QY 241 TGGCCACCCGCTGCGAGAGGACATGAGATGACAAAGCTCAGAGAAAGCCCTGAG 300
Db 297 TGGCCACCCGCTGCGAGAGGACATGAGATGACAAAGCTCAGAGAAAGCCCTGAG 356
QY 301 GTTCTTACATCAAAAGAAAGAGGTGTCCAAACGGTGTGTCAGCAAGACTTACAGAGA 360
Db 357 GTTCTTACATCAAAAGAAAGAGGTGTCCAAACGGTGTGTCAGCAAGACTTACAGAGA 416
QY 361 GAGGGGAGTGGAGCCACTCAGCCACAGGTACGAGAGGATGCTGTGTGTTGAACCT 420
Db 417 GAGGGGAGTGGAGCCACTCAGCCACAGGTACGAGAGGATGCTGTGTGTTGAACCT 476
QY 421 GGGCAGCCAGAGATGACATTGACAGAACTCTCCACAGCCAGCGTCCCAACGGGAGG 480
Db 477 GGGCAGCCAGAGATGACATTGACAGAACTCTCCACAGCCAGCGTCCCAACGGGAGG 536
QY 481 AGAAATGTGCGAATCTGTGTCTGAGCTTAAACAGGGCTGGAGAGTGTAGAGCAGAG 540
Db 537 AGAAATGTGCGAATCTGTGTCTGAGCTTAAACAGGGCTGGAGAGTGTAGAGCAGAG 596
QY 541 GAGCCACATGAGGAGTGCAGCGTAGACAGAGCAGCGGTCTATGGAGGAGAGCT 600
Db 597 GAGCCACATGAGGAGTGCAGCGTAGACAGAGCAGCGGTCTATGGAGGAGAGCT 656
QY 601 GAGGAGGCGCCAGCAGGATGAGTGCAGGTGGTGTGTCAGGATCAAGCGCCCTTG 660
Db 657 GAGGAGGCGCCAGCAGGATGAGTGCAGGTGGTGTGTCAGGATCAAGCGCCCTTG 716
QY 661 CCTCCAGGTAAACAGATTACAGAGAACTCACTGCAAGACCCCAACAGAAATATAGC 720
Db 717 CCTCCAGGTAAACAGATTACAGAGAACTCACTGCAAGACCCCAACAGAAATATAGC 776
QY 721 CCAGTGGCAACTTGAAGGAGATGGCAGCAGTGGCTGATCAACACATACATCCCAAG 780
Db 777 CCAGTGGCAACTTGAAGGAGATGGCAGCAGTGGCTGATCAACACATACATCCCAAG 836
QY 781 AAGCTCAATCTTTCAGTGAAGTTTGATTACGAGCTGGCATGTCACCCGCTACAC 840
Db 837 AAGCTCAATCTTTCAGTGAAGTTTGATTACGAGCTGGCATGTCACCCGCTACAC 896
QY 841 AAAGGAGATGAGGCTATGCGCCGCCCAAGAGGAAACCAAACTGCTGAAAGGGCCAG 900
Db 897 AAAGGAGATGAGGCTATGCGCCGCCCAAGAGGAAACCAAACTGCTGAAAGGGCCAG 956
QY 901 CGTGTGAGAGCACATCTACAGGAAATGATGGAATGCTGCTTATTTATCTGCAATG 960
Db 957 CGTGTGAGAGCACATCTACAGGAAATGATGGAATGCTGCTTATTTATCTGCAATG 1016
QY 961 GCTCGCCACAGAGAGATGSCAAGATCCAGGTTACTTTTGGAGATCTCTTTCAGAGATAC 1020
Db 1017 GCTCGCCACAGAGAGATGSCAAGATCCAGGTTACTTTTGGAGATCTCTTTCAGAGATAC 1076
QY 1021 GTTGTGATTTACATAAAGTGTGGCATCTCTCATCGTCCAGGAAACATGAGTGTGTA 1080
Db 1077 GTTGTGATTTACATAAAGTGTGGCATCTCTCATCGTCCAGGAAACATGAGTGTGTA 1136
QY 1081 GACTTTGAGGAGAGATGCTATCGCAAGCCGAGATGACCATCTTGTGATTAGCTACTC 1140
Db 1137 GACTTTGAGGAGAGATGCTATCGCAAGCCGAGATGACCATCTTGTGATTAGCTACTC 1196

QY 1141 AAGTGA 1146
Db 1197 AAGTGA 1202

RESULT 3
AK092694

LOCUS Homo sapiens cDNA FLJ35375 fis, clone SKMUS2002634. 1322 bp mRNA linear PRI 15-JUL-2002

DEFINITION AK092694

ACCESSION AK092694

VERSION AK092694.1 GI:21751348

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Tashiro, H., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Makamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Makamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Watanabe, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.

TITLE NEDO human cDNA sequencing project

REFERENCE 2 (bases 1 to 1322)

AUTHORS Isogai, T. and Yamamoto, J.

TITLE Direct Submission

COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' - & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES

source

1..1322

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="SKMUS2002634"

/tissue_type="skeletal muscle"

/clone_lib="SKMUS2"

/note="Cloning vector: pME18SFL3"

57..1202

/note="unnamed protein product"

/codon_start=1

/protein_id="BAC03948.1"

/db_xref="GI:21751349"

/translation="MARGKEGEGEPKALRKTATLVIILARGHWOANENSIFQACEFTGMPGTQSPQAPKPTPTTSHQKQAPKPPRLPHGHGCGQSEKAPVSHIKKVEKTVKTVSRGDSVSHLSHRYERDAGLVLEPGQFENDIDRLSHSGSPTRRKCANLVSELTGKRWVMEQBEPTVRSVSDTSDSGYGEAEPEEPQGVQVAVRIKRLPSQVNRFTKLNCKAQKYSPVGNLKGWQWADHIOQKLNPFSEEDVELAMSTRLHGDEGYRKEGKTAEAKRAEHIYREMDMCFITCTMARHRRDKIQVTFGLDLPYRVISDKVVGILMRARKHGLVDFEGEMLWQGRDDHVVITLLK"

CDS

Query Match 99.9%; Score 1144.4; DB 9; Length 1322;
Best Local Similarity 99.9%; Pred. No. 1.5e-304;
Matches 1145; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

ORIGIN

1 ATGGCTCCGGCGGAAAGGAGTGGGAGGGCCCAAGCAAGAGCGCCCTCCGGAAGATA 60
57 ATGGCTCCGGCGGAAAGGAGTGGGAGGGCCCAAGCAAGAGCGCCCTCCGGAAGATA 116

Db	716	CCCTCCACGGTAACAGATTTTACAGAGAAACTCAACTGCAGGCCCAACAGAAATATAGC	775
Qy	721	CCAGTGGGCAACTTTGAAAGGGAGATGCGACAGCTGGGCTGATGAAACACATACAAATCCCAG	780
Db	776	CCAGTGGGCAACTTTGAAAGGGAGATGCGACAGCTGGGCTGATGAAACACATACAAATCCCAG	835
Qy	781	AAGCTCAATCCTTTTCAGTGAAGAGTTTGATTACGAGCTGSCCAATGTCACCCGCCCTACAC	840
Db	836	AAGCTCAATCCTTTTCAGTGAAGAGTTTGATTACGAGCTGSCCAATGTCACCCGCCCTACAC	895
Qy	841	AAAGGAGATGAGGGCTATGCGCGCCCCCAAGAGAGGAACCAAAACTGCTGAAGGGGCCAAG	900
Db	896	AAAGGAGATGAGGGCTATGCGCGCCCCCAAGAGAGGAACCAAAACTGCTGAAGGGGCCAAG	955
Qy	901	CGTGCTGAGAGACACATCTACAGGAAATGATGGACATGTCGCTTATTAATCTGCACAAATG	960
Db	956	CGTGCTGAGAGACACATCTACAGGAAATGATGGACATGTCGCTTATTAATCTGCACAAATG	1015
Qy	961	GCTCGCCACACAGACGAGATGGCAAGATCCAGGTTACTTTTGGAGATCTCTTTGACAGATAC	1020
Db	1016	GCTCGCCACACAGACGAGATGGCAAGATCCAGGTTACTTTTGGAGATCTCTTTGACAGATAC	1075
Qy	1021	GTTCGTATTTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCAGGAACCATGACCTGGTA	1080
Db	1076	GTTCGTATTTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCAGGAACCATGACCTGGTA	1135
Qy	1081	GACPTTCAAGGAGAGATGCTATGCAAGGCCGAGATGACCATGTGTGTGATTACGCTACTC	1140
Db	1136	GACPTTCAAGGAGAGATGCTATGCAAGGCCGAGATGACCATGTGTGTGATTACGCTACTC	1195
Qy	1141	AAGTGA	1146
Db	1196	AAGTGA	1201

RESULT 6	AF504061	1128 bp	linear	ROD 01-JUL-2002
LOCUS	AF504061			
DEFINITION	Mus musculus striated muscle activator of Rho-dependent signaling mRNA, complete cds.			

AF504061.
 AF504061.1 GI:20805880
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1126)
 Arai, A., Spencer, J.A. and Olson, E.N.
 STARS, a Striated Muscle Activator of Rho Signaling and Serum
 Response Factor-dependent Transcription
 J. Biol. Chem. 277 (27), 24453-24459 (2002)
 JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 1128)
Arai, A., Spencer, J.A. and Olson, E.N.
Direct Submission
Submitted (19-APR-2002) Molecular Biology, University of Texas
Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX
75390-9148, USA

FEATURES	source	Location/Qualifiers
		1..1128
		/organism="Mus musculus"
		/mol_type="mRNA"
		/db_xref="taxon:10090"
CDS		1..1128
		/note="STARS"
		/codon_start=1
		/product="striated muscle activator of Rho-dependent signaling"
		/protein_id="AA388877.1"
		/db_xref="GI:20805881"
		/translation="MAGGERERAGPAKSALRKVKTATLVNLARGQQQWANEINSTKQ

ORIGIN		Query Match	69.0%;	Score 791.2;	DB 10;	Length 1128;
		Best Local Similarity	81.6%;	Pred. No. 5e-207;		
Matches	Conservative	935;	0;	Mismatches 193;	Indels 18;	Gaps 1;
Qy	1	ATGCGTCCGGCGAAAGAAAGAAAGCGGGAGGGCCAGCCAAAGAGCGCCCTCCCGAAGATA	60			
Db	1	ATGCGTCCAGGAGAAAGAAAGGAGCGGGCCCGGCCAAGAGTGCCTCCCGAAGGTC	60			
Qy	61	CGCAGCCACCTGTGTATCAGCTTTGGCCGAGGTTGGCAGCAGTGGGCGAATAGAAAC	120			
Db	61	CGCAGCAGCAACCTGGTTATCAATTTGGCCGAGGTTGGCAGCAGTGGGCGAATAGAAAC	120			
Qy	121	AGCATCAGCAGCGCCAGGAGCCTACAGGCTGCTGCCGGAGGGACCCAGGACTCACCT	180			
Db	121	AGTACCAACAGGCCCCAGGAGCCTGCAGGCTGCTGCCGGAGCAACTCATGACGTACCT	180			
Qy	181	CAAGCTCCTAAACCAATCAACCCCCCTACTTCAACACGAAAGACTCAGAGTGCCCCAAAG	240			
Db	181	AACGCTCTTAAAGAGCGCGTCC-----TTACCAGCATGCCCCCAA 222				
Qy	241	TCGCCACCCCGCTCGCAGNAGGACATGGAGATGGCAAGACTCAGAGAAGCCCTGAG	300			
Db	223	ACTCTGTCTCAAAGCCAGATCGAGACGGAGAGGGHCAACACTCAGAAAGGCCACCGAG	282			
Qy	301	GTTTCTCACATCAAAAGAAAGAGGTGTCCAAACCGTGGTTCAGCAAGACTTACGAGAGA	360			
Db	283	GTCTCCACATTAAGGAAAGAGGTGACCCAGAACGGTTGTTCAGCAAGGCTTATGAGAGG	342			
Qy	361	GGAGGGACGTGAGCCACCTCAGCCACAGGTACGAGAGGATGCTGGTGTGCTTGAACCT	420			
Db	343	GGAGGAGATGTGAACCTACCTGAGCCACAGGTATGAGAATGATGGTGGCGTGTCTGAAGCT	402			
Qy	421	GGCAGCCAGAGAAATGACATTGACAGAACTCCTCCACAGCCACGGCTCCCCAAACGCGGAGG	480			
Db	403	ATTTCAGCCAGAGAAATGACATTGACAGAACTCTTCTTAGTCAGACTCGCCCAACACGGAGA	462			
Qy	481	AGAAATGTGCCAACCTGTGTCTGAGCTTAACCAAGGCTGGAGAGTGTATGATGAGCAGGAG	540			
Db	463	AGAAATGCAACCAACCTGTGTCTGAGCTCACCAGAGGCTTGGAAAGTGTATGGAACAGGAA	522			
Qy	541	GAGCCACATGGAGGAGTGACAGGTAGACACAGAGGACACGGCTATGAGGAGAGAGGCT	600			
Db	523	GAGCCACGTGGAGAGTGAACGGTATAGACAGAGGACAGTGGCTACGAGGGGATATG	582			
Qy	601	GAGGAGAGGCCCGAGCAGGATGGAGTGCAGGTGGCTGTGTTCAGGATCAAGCGCCCTTG	660			
Db	583	GAGGAGAGGCCCTGAGCAAGATGCAGCGCTGTGGCTCCTGCCAGGATCAACCGCCCTTG	642			
Qy	661	CCCTCCAGGTAAACAGATTACAGAGAACTCAACTGCAAGCCCAACACAGAAATATAGC	720			
Db	643	CACCTCCAGGCAACAGGTACTCTGAGCCCACTCAACTGTAAAGGCCCATCGAAATACAGC	702			
Qy	721	CCAGTGGGCAACTTGAAGGGGAGATGGCAGCAGTGGGCTGATGAACACATACATCCCAAG	780			
Db	703	CAAGTGACAACTTGAAGGGAGGTGCGCAGCGTGGCCGATGAACACGCTCCAGTCCAG	762			
Qy	781	AAGCTCAATCGTTTCAGTGAAGATTTTGANTACGAGCTGGCCATGTCTCCACCCGCTACAC	840			
Db	763	AAGCTCAATCCCTTCAGTGAAGATTTTGAATATGACCTAGCCATGTCCACTCGGCTCCAC	822			
Qy	841	AAAGGAGATGAGGGCTATGGCCGCCCCCAAGAGGAACCAAACTGTCTGAAGAGGCCCAAG	900			
Db	823	AAGGAGAGCAGGGCTATGGCCGCCCCCAAGAGGAAGCAAGACAGCTGAAAGGGCCCAAG	882			
Qy	901	CGTCTCAGGAGGACATCTTACAGGGAATGATGGACATGTGCTTATTATCTGCAACAATG	960			

```

Db      883  CGAGCGGAGACCATCTATCGGGAATATTATGGAACTGTCTTTGTTATCCGACATG 942
Qy      961  GCTCCGCACAGACGAGATGGCAAGATCCAGGTTACTTTTGAGATCTCTTTTGACAGATAC 1020
Db      943  GCTGCCACAGACGAGATGGCAAGATCCAGGTTACTTTTCGAGAACTCTTTTGATCGCTAT 1002
Qy      1021 GTTCGTATTTCAGATAAAGTAGTGGCCATCTTCATGCGTGCCAGGAAACATGGACTGGTA 1080
Db      1003 GTTCGCATTTCTGATAAAGTCGTGGGCATCTCTATGCGTGCCAGGAAACAGGACTGGT 1062
Qy      1081 GACTTTGAAGGAGAGATGCTATGGAAGCCGAGATACCATTGTTGATTAACGCTACTC 1140
Db      1063 CACTTTGAAGGAGAGATGCTATGGAAGCCGAGACGACCATGTTGTGATTACTCTCGT 1122
Qy      1141 AAGTGA 1146
Db      1123 GAGTAA 1128

RESULT 7
AF336113
LOCUS      AF336113                1285 bp      mRNA      linear      ROD 13-AUG-2002
DEFINITION Rattus norvegicus MS1 mRNA, complete cds.
ACCESSION  AF336113
VERSION     AF336113.1  GI:22212591
KEYWORDS
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
REFERENCE  1 (bases 1 to 1285)
AUTHORS   Mahadeva,H., Brooks,G., Lodwick,D., Chong,N.W. and Samani,N.J.
TITLE      MS1, a novel stress-responsive, muscle-specific gene that is
           up-regulated in the early stages of pressure overload-induced left
           ventricular hypertrophy
JOURNAL    FEBS Letters (2002) In press
REFERENCE  2 (bases 1 to 1285)
AUTHORS   Mahadeva,H. and Samani,N.J.
TITLE      Direct Submission
JOURNAL    Submitted (14-JAN-2001) Department of Cardiology, Leicester
           University, Groby Road, Leicester LE3 9QP, England
FEATURES   Location/Qualifiers
           1..1285
           /organism="Rattus norvegicus"
           /mol_type="mRNA"
           /strain="WKY"
           /db_xref="taxon:10116"
           54..1181
           /codon_start=1
           /product="MS1"
           /protein_id="AA094370.1"
           /db_xref="GI:22212592"
           /translation="MAPGETVREAGPAKSLQKVRRLTLVNLARGWQWANEINSTRQ
           ADEPAGLPGATQDLPHPKPEPRQHPKPPKPDGDRGSEEADEVSHIMKRE
           VRTVVSXAYERGSDVNYLSHRYBHDGVSVAOPNDIDRILSHDSPTRRKCTNL
           VSKLTKGWKVPQSEPKWSDSIDTDSGYGDMBERPEQDVQVAARIAIKPLHSQA
           NPYSLTNLKAHKYKYSQVDNLKRWQWQADEHIQSKLNPSPDEFDYLAMSTRLHKQ
           DPGYRKGESKTAERAKAAEHIITREIMELCFVIRTPARHRDQKIQVTFGELFDRI
           VRISDKVGIIMRARRKHGLVHFEGLMWGLQKDDHVVITLLE"
CDS
Query Match      68.5%; Score 784.8; DB 10; Length 1285;
Best Local Similarity 81.2%; Pred. No. 2.9e-205;
Matches 931; Conservative 0; Mismatches 197; Indels 18; Gaps 1;
Qy      1  ATGGCTCCGGCGGAAAGAAAGCGGGAGGGCCCAAGCCCAAGAGCGCCCTCCGGAAGATA 60
Db      54  ATGGCTCTCGGAGAACTGTAGGAGAGCGGGCGCGCCCAAGAGTGCCTCCCAAGAGTGC 113
Qy      61  CGCACAGCCACCTGCTCATACGTTGGCCCGGAGTTTGCACAGTGGCGGATGAGAAC 120

```

```

114  CGCAGAGCAACCTCGTGTGATCACTTGGCCCGAGGATGCGACGAGTGGCGAATGAGAAC 173
121  AGCATCAGGAGGCCCGCAGAGAGCTACAGGCTGGCTGCGCGGAGGAGCCAGGACTCACCT 180
174  AGTACCAGACAGGCCCGAGGAGCCTGAGGCTGGCTGCCAGGAGCAACTCAAGACCTTACCT 233
181  CAAGCTCTCTAAACCAATACACACCCCTTACTTACACACGAAAGCTCAGAGTGGCCCAAG 240
234  CACACTCTCTAAAGAACCGGGTCTC-----GCCAGCATGCCCCCAA 275
241  TCGCACCCCGCTCGCAGNAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCTGAG 300
276  CCTCCGCTCTCAAGCCAGATGGAGACAGAGAGGGGACGAGGCTCTGAGGAGGACCGGAG 335
301  GTTCTTCACATCAAAAGAAAGAGGTGTCCAAACCGTGTGTCAGCAAGACTTACGAGAGA 360
336  GTCTCCCATATCAAAAGGAAAGAGGTGACCAAGACGGTTGTCAETAAAGGGTATGAGAGA 395
361  GGAGGGGACGTGAGCCACCTCAGCCACAGGTACGAGAGGATGCTGGTGTGTTGAACCT 420
396  GAGGAGCATGTGAATACCTGAGCCACAGGTACGAGCATGATGGCGGCGTGTCTGAAGCG 455
421  GGCAGCCAGAGATGACATTGACAGAACTCTCCACAGCCACGCTCCCAACCGCGAGG 480
456  GTCAGCCAGACATGACATTGACAGATCTCTCTCAGTCACGACTCGCCACGCGGAGA 515
481  AGAAATGTGCCAACCTGGTGTCTGAGCTAAACCAAGGGCTGGAGAGTATGAGAGCAGGAG 540
516  AGAAATGCAACCAACTGGTATCTAAGCTGACCAAGGCTGGAAAGTATGATGGAACAGGAG 575
541  GAGCCACATGGAGGAGTGACAGCTAGACACAGAGGACAGCGCTATGGAGAGAGGCT 600
576  GAGCCCAAGTGGAGAGTGACAGCATAGACACAGAGACAGTGGTACGAGGGGACATG 635
601  GAGGAGAGGCCCGCAGCAGGATGGAGTGCAGTGGCTGTGTTGAGTCAAGCCGCCCTTG 660
636  GAGGAGAGGCTTGAGCAAGATGTAGCGCAGTGGCTGCTGCCAGGATTAACGCCCTTG 695
661  CCCTCCAGGTAAACAGATTACAGAAACTCACTGTCGAAGCCCAACAGAAATATAGC 720
696  CACTCCAGGCAACACAGATACTCAGAGACACTCAACTGTAAAGGCCCATCGGAATACAGC 755
721  CCAGTGGGCAACTTGAAGGGAGATGGCAGCAGTGGGCTGATGAACACATACATCCAG 780
756  CAAGTGGACAACTTGAAGGCGAGTGGCAGCAGTGGGCCGATGAACATACACTACAG 815
781  AAGCTCAATCTTTCAGTGAAGATTTGATTACAGCTGGCCATGTGCCACCCGCTACAC 840
816  AAGCTCAACCCCTTCAGTGATGAATTTGACTTATGACCTAGCCATGTCCACTCGACTCCAC 875
841  AAGGAGATGAGGGCTATGGCCGCCCAAGAGGAAACCAAACTGCTGAAAGGGCCCAAG 900
876  AAGGAGACAGGGATATGGCCGCCCAAGAGGAAAGCAAGACAGCTGAAGGGCCCAAG 935
901  CGTCTCAGGAGACATCTTACAGGAAATGATGGACATGTGCTTCAATTTCTGCACAATG 960
936  AGAGCCGAGGAGACATCTATCGGAAATTTATGGAATTTGCTTTGTTATCCGCACAATG 995
961  GCTCGCACAGACGAGATGGCAGATCCAGGTTACTTTTGAGATCTCTTTTGACAGATAC 1020
996  GCTCGCCATAGACAGATGGCAGATCCAGGTACTCTTCGGAGAACTCTTTTGTAGATAT 1055
1021 GTTCGTATTTTCAGATAAAGTAGTGGGCATTTCTCATGCGTGGCCAGGAAACATGAGCTGGTA 1080
1056 GTTCGAATTTTCAGATAAAGTGGTGGGCATCTCTCATGCGTGGCCAGGAAACACGAGCTGGT 1115
1081 GACTTTGAAGGAGAGATGCTATGGAAGCCGAGATGACCATGTTGTTGATTACGCTACTC 1140
1116 CACTTTGAAGGGGAGATGCTATGGAAGGCAAGATGACCATGTTGTGATTACTTCTCCTT 1175
1141 AAGTGA 1146
1176 GAGTAA 1181

```

```

RESULT 8
AP000856/c      152130 bp   DNA      linear   PRI 01-MAR-2001
LOCUS           Homo sapiens genomic DNA, chromosome 8q23, clone:XB1539E1.
DEFINITION      AP000856
ACCESSION       AP000856
VERSION         AP000856.3 GI:13358956
KEYWORDS        Homo sapiens (human)
SOURCE          Homo sapiens
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1 (sites)
AUTHORS         Shimizu,N. and Asakawa,S.
TITLE           Homo sapiens DNA chromosome 8 SEQUENCE
JOURNAL         Published Only in DataBase (1999)
REFERENCE       2 (bases 1 to 152130)
AUTHORS         Shimizu,N. and Asakawa,S.
TITLE           Direct Submission
JOURNAL         Submitted (13-DEC-1999) Nobuyoshi Shimizu, Keio University, School
                of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
                160-8582, Japan (E-mail:nshimizu@med.keio.ac.jp,
                Tel:81-3-3351-2370, Fax:81-3-3351-2370)
COMMENT         On Mar 16, 2001 this sequence version replaced gi:8096485.
FEATURES        Location/Qualifiers
                 source          1..152130
                                /organism="Homo sapiens"
                                /mol_type="genomic DNA"
                                /db_xref="taxon:9606"
                                /chromosome="8"
                                /map="8q23"
                                /clone="KB1539E1"
                                /cell_line="FLEB 14 - 14"
                                /clonc_lib="Keio BAC library"
repeat_region   1603..1809
                                /evidence=not experimental
                                /rpt_family="AluJb"
repeat_region   1819..2094
                                /evidence=not experimental
                                /rpt_family="AluSg1"
repeat_region   2098..2136
                                /evidence=not experimental
                                /rpt_family="(CAA)n"
repeat_region   complement(3276..3445)
                                /evidence=not experimental
                                /rpt_family="MIR"
repeat_region   3528..3670
                                /evidence=not experimental
                                /rpt_family="L1MC5"
repeat_region   3737..3827
                                /evidence=not experimental
                                /rpt_family="MERSA"
repeat_region   3888..4246
                                /evidence=not experimental
                                /rpt_family="L1MC5"
                                /complement(4307..4612)
repeat_region   /evidence=not experimental
                                /rpt_family="AluSg"
                                /complement(6251..6622)
repeat_region   /evidence=not experimental
                                /rpt_family="ITR37A"
                                /complement(6627..6929)
repeat_region   /evidence=not experimental
                                /rpt_family="AluY"
                                /complement(7278..7453)
repeat_region   /evidence=not experimental
                                /rpt_family="MERSA"
                                /complement(7484..7774)
repeat_region   /evidence=not experimental
                                /rpt_family="AluSg"
                                /complement(7792..7991)
repeat_region   /evidence=not experimental

```

```

repeat_region   /rpt_family="LTR54"
                                /complement(8094..8190)
                                /evidence=not experimental
                                /rpt_family="LTR54"
repeat_region   /complement(8417..8474)
                                /evidence=not experimental
                                /rpt_family="MER110"
repeat_region   9119..9260
                                /evidence=not experimental
                                /rpt_family="(TA)n"
repeat_region   /complement(10809..10935)
                                /evidence=not experimental
                                /rpt_family="FLAM A"
repeat_region   /complement(11155..11575)
                                /evidence=not experimental
                                /rpt_family="MLT1J1"
repeat_region   /complement(12042..12257)
                                /evidence=not experimental
                                /rpt_family="MIR"
repeat_region   12298..12319
                                /evidence=not experimental
                                /rpt_family="AT rich"
                                /complement(12382..12476)
repeat_region   /evidence=not experimental
                                /rpt_family="MIR"
repeat_region   12605..12686
                                /evidence=not experimental
                                /rpt_family="MIR"
                                /complement(13249..13444)
repeat_region   /evidence=not experimental
                                /rpt_family="MLT1I"
repeat_region   13704..13811
                                /evidence=not experimental
                                /rpt_family="MERSA"
                                /complement(13841..13951)
repeat_region   /evidence=not experimental
                                /rpt_family="L2"
repeat_region   /complement(14971..15272)
                                /evidence=not experimental
                                /rpt_family="AluSx"
repeat_region   /complement(16581..16677)
                                /evidence=not experimental
                                /rpt_family="AluJo"
repeat_region   17788..18078
                                /evidence=not experimental
                                /rpt_family="AluJb"
repeat_region   18079..18098
                                /evidence=not experimental
                                /rpt_family="(CAA)n"
repeat_region   18556..18577
                                /evidence=not experimental
                                /rpt_family="AT rich"
repeat_region   /complement(18578..18860)
                                /evidence=not experimental
                                /rpt_family="AluY"
repeat_region   18861..18882
                                /evidence=not experimental
                                /rpt_family="AT rich"
repeat_region   /complement(19014..19214)
                                /evidence=not experimental
                                /rpt_family="MIR"
repeat_region   /complement(19652..20150)
                                /evidence=not experimental
                                /rpt_family="L1MB5"
repeat_region   /complement(20157..20921)
                                /evidence=not experimental
                                /rpt_family="L1M4"
                                /complement(20936..21000)
repeat_region   /evidence=not experimental
                                /rpt_family="MER91B"
repeat_region   21203..23651
                                /evidence=not experimental
                                /rpt_family="L1M4"

```

```

repeat_region 23654..24174
/evidence=not experimental
/rpt_family="LIMB8"
complement(24671..24701)
/evidence=not experimental
/rpt_family="L2"
24911..24932
/evidence=not experimental
/rpt_family="AT rich"
complement(25053..25411)
/evidence=not experimental
/rpt_family="MLT1A1"
25476..25577
/evidence=not experimental
/rpt_family="MER5A"
complement(25502..25607)
/evidence=not experimental
/rpt_family="MER5A"
25685..25705
/evidence=not experimental
/rpt_family="AT rich"
complement(28001..28123)
/evidence=not experimental
/rpt_family="FLAM_C"
28582..28684
/evidence=not experimental
/rpt_family="FLAM_A"
28892..28972
/evidence=not experimental
/rpt_family="MIR"
29596..29893
/evidence=not experimental
/rpt_family="AluSx"
29894..30069
/evidence=not experimental
/rpt_family="AluSg/x"
31207..31232
/evidence=not experimental
/rpt_family="AT rich"
complement(31446..31520)
/evidence=not experimental
/rpt_family="MIR"
31705..31746
/evidence=not experimental
/rpt_family="ATGn"
31775..31993
/evidence=not experimental
/rpt_family="MER46A"
complement(32042..32122)
/evidence=not experimental
/rpt_family="MIR"
32464..32763
/evidence=not experimental
/rpt_family="MER1B"
complement(33126..33290)
/evidence=not experimental
/rpt_family="MER5A"
complement(33707..34067)

Query Match 58.2%; Score 667.4; DB 9; Length 152130;
Best Local Similarity 99.9%; Pred. No. 7,7e-173;
Matches 668; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGCTCCGGGCGAAAGAAAGCGGGAGGGCCAGCAAGAGCGCCCTCCGGAAGATA 60
Db 139073 ATGGCTCCGGGCGAAAGAAAGCGGGAGGGCCAGCAAGAGCGCCCTCCGGAAGATA 139014

QY 61 CGCACAGCCACCTGTGTATCAGTTGGCCCGAGGTGGCAGAGTGGCGGCAATGAGAAC 120
Db 139013 CGCACAGCCACCTGTGTATCAGTTGGCCCGAGGTGGCAGAGTGGCGGCAATGAGAAC 138954

QY 121 AGCATCAGGAGCCAGGAGCGCTACAGGTGGCTGCCGGGGAGCCAGGACTCACCT 180

```

```

Db 138953 AGCATCAGGAGCGCCAGGAGCGCTACAGGTGGCTGCCGGAGGGAGCCAGGACTCACCT 138894
QY 181 CRAGCTCCTAAACCAATCACACCCCTACTTACACACAGAAAGCTCAGAGTGCCCAAG 240
Db 138893 CAAGCTCCTAAACCAATCACACCCCTACTTACACACAGAAAGCTCAGAGTGCCCAAG 138834
QY 241 TCGCCACCCCGCTGCCAGAAAGCATGGAGATGGACAAAGCTCAGAGAAAGCCCTGAG 300
Db 138833 TCGCCACCCCGCTGCCAGAAAGCATGGAGATGGACAAAGCTCAGAGAAAGCCCTGAG 138774
QY 301 GTTCTTCACATCAAAAAGAAAGAGGTGTCCAAAACGGTGTTCAGCAAGACTTACGAGAGA 360
Db 138773 GTTCTTCACATCAAAAAGAAAGAGGTGTCCAAAACGGTGTTCAGCAAGACTTACGAGAGA 138714
QY 361 GGAGGGAGCTGAGCCACCTCAGCCACAGATACGAGAGGAGATCTGGTGTGCTTGAACCT 420
Db 138713 GGAGGGAGCTGAGCCACCTCAGCCACAGATACGAGAGGAGATCTGGTGTGCTTGAACCT 138654
QY 421 GGGCAGCCAGAGATGACATTGACAGATCTCCACAGCCACCGCTCCCCAACCGCGAGG 480
Db 138653 GGGCAGCCAGAGATGACATTGACAGATCTCCACAGCCACCGCTCCCCAACCGCGAGG 138594
QY 481 AGAAATGTGCCAACTGTGTCTGAGCTAAACCAAGGGCTGGAGAGTGTATGAGCAGGAG 540
Db 138593 AGAAATGTGCCAACTGTGTCTGAGCTAAACCAAGGGCTGGAGAGTGTATGAGCAGGAG 138534
QY 541 GAGCCACATGGAGGAGTGACAGCTAGACACAGAGAGCAGCGGCTATGAGAGGAGGCT 600
Db 138533 GAGCCACATGGAGGAGTGACAGCTAGACACAGAGAGCAGCGGCTATGAGAGGAGGCT 138474
QY 601 GAGGAGAGGGCCGAGCAGGATGGAGTGGCTGTGGTCAAGATCAAGCGCCCTTG 660
Db 138473 GAGGAGAGGGCCGAGCAGGATGGAGTGGCTGTGGTCAAGATCAAGCGCCCTTG 138414
QY 661 CCTCCCGAG 669
Db 138413 CCTCCCGAG 138405

RESULT 9
AF303733/c
LOCUS
DEFINITION Homo sapiens chromosome 8 clone RP11-79F7 map 8q23, WORKING DRAFT
SEQUENCE, 10 unordered pieces.
ACCESSION AF303733
VERSION AF303733.2 GI:14389401
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_CANCELLED.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 153147)
Reichwald, K., Baumgart, C., Blechschmidt, K., Dette, M., Jahn, N.,
Lehmann, R., Menzel, U., Pooley, A., Schilhabel, M.B., Schudy, A.,
Siddiqui, R., Taudien, S., Wen, G., Rosenthal, A. and Platzer, M.
Chromosome 8 genomic sequence
Unpublished
2 (bases 1 to 153147)
Genome Sequencing Center Jena.
Direct Submission
Submitted (05-SEP-2000) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
On Jun 13, 2001 this sequence version replaced gi:10442740.
COMMENT
Center: Institute of Molecular Biotechnology
Center code: IMB
Web site: http://genome.imb-jena.de/
Contact: gscj-submit@genome.imb-jena.de
----- Project Information
Center project name: H552
Center clone name: RP11-79F7
----- Summary Statistics
Sequencing vector: M13; 100% of reads

```


REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 5 (bases 1 to 186751)
 Waterston,K.
 Direct Submission
 Submitted (09-MAY-2001) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Nov 8, 2000 this sequence version replaced gi:7630974.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@watson.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0395G23

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., rateno,M., Catanesi,J.J. and de Jong,P.J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-649G15. Actual start of this clone is at base position 1 of RP11-395G23; actual end is at base position 186751 of RP11-395G23.

There are polymorphic base pair differences in the overlap between the clone RP11-395G23 and RP11-649G15.

FEATURES

source

Location/Qualifiers
 1. .186751
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="8"
 /map="8"
 /clone="RP11-395G23"
 /clone_lib="RP11-11"
 90. .192
 /rpt_family="Alu"
 400. .514
 /rpt_family="MIR"
 1104. .1401
 /rpt_family="Alu"
 1402. .1577
 /rpt_family="Alu"
 2954. .3028
 /rpt_family="MIR"
 3283. .3501
 /rpt_family="MER2_type"

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region
 3550. .3630
 /rpt_family="MIR"
 repeat_region
 3372. .4271
 /rpt_family="MER1_type"
 misc_feature
 4255. .4618
 /note="similar to EST AA347139 (NID:G1999427)"
 repeat_region
 4634. .4798
 /rpt_family="MER1_type"
 repeat_region
 5215. .5819
 /rpt_family="MER2_type"
 misc_feature
 6093. .6323
 /note="similar to EST A1419746 (NID:G4265677) t951f06.x1"
 repeat_region
 6631. .6760
 /rpt_family="MER1_type"
 repeat_region
 6789. .6815
 /rpt_family="MER1_type"
 repeat_region
 6816. .7300
 /rpt_family="ERV1"
 repeat_region
 7301. .7483
 /rpt_family="MER1_type"
 repeat_region
 7488. .7657
 /rpt_family="MER1_type"
 repeat_region
 8138. .8260
 /rpt_family="L1"
 repeat_region
 8278. .8544
 /rpt_family="L1"
 repeat_region
 8640. .8787
 /rpt_family="L1"
 repeat_region
 8855. .9001
 /rpt_family="MaLR"
 repeat_region
 9287. .9377
 /rpt_family="MER1_type"
 repeat_region
 9684. .9819
 /rpt_family="L2"
 repeat_region
 10686. .10733
 /rpt_family="L1"
 misc_feature
 10725. .10947
 /note="similar to EST H65488 (NID:G1024228) yr56a02.s1"
 repeat_region
 10928. .11231
 /rpt_family="L1"
 repeat_region
 11232. .11447
 /rpt_family="Alu"
 repeat_region
 11448. .11518
 /rpt_family="L1"
 repeat_region
 11596. .11697
 /rpt_family="L1"
 repeat_region
 11899. .12354
 /rpt_family="L1"
 repeat_region
 12581. .12897
 /rpt_family="Alu"
 repeat_region
 13010. .13386
 /rpt_family="MaLR"
 repeat_region
 13387. .14996
 /rpt_family="MaLR"
 repeat_region
 14997. .15365
 /rpt_family="MaLR"
 repeat_region
 15470. .15708
 /rpt_family="L1"
 repeat_region
 15736. .15807
 /rpt_family="L1"
 repeat_region
 15820. .17099
 /rpt_family="L1"
 repeat_region
 17820. .18124
 /rpt_family="L1"
 repeat_region
 18314. .18470
 /rpt_family="L2"
 misc_feature
 19602. .19688
 /note="similar to EST AW504109 (NID:G7141776)"
 misc_feature
 19605. .19688
 /note="similar to EST AW918394 (NID:G8084170)"
 misc_feature
 19606. .19686
 /note="similar to EST AW318757 (NID:G6748301) un05h10.y1"
 repeat_region
 21909. .22220

```
repeat_region /rpt_family="Alu"
22662..22964
misc_feature /rpt_family="Alu"
23594..23647
misc_feature /note="similar to EST AW318757 (NID:96748301) un05h10.y1"
23596..23703
misc_feature /note="similar to EST AW504109 (NID:97141776)"
23596..23633
misc_feature /note="similar to EST AW918394 (NID:98084170)"
24163..24399
repeat_region /rpt_family="L1"
24643..24756
misc_feature /note="similar to EST AW504109 (NID:97141776)"
25651..25920
misc_feature /note="similar to EST AA367453 (NID:92019771)"
26114..26674
repeat_region /rpt_family="L1"
27351..27417
repeat_region /rpt_family="MIR"
27422..27740
repeat_region /rpt_family="ERV1"
27785..27836
repeat_region /rpt_family="ERV1"
27837..28123
repeat_region /rpt_family="Alu"

Query Match 58.2%; Score 667.4; DB 9; Length 186751;
Best Local Similarity 99.9%; Pred. No. 7.7e-173;
Matches 669; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCTCCGGGCGGAAAGGAGGCGGAGGCGCCAGGAGGCGCCCTCCGAGAGATA 60
Db 110228 ATGCTCCGGGCGGAAAGGAGGCGGAGGCGCCAGGAGGCGCCCTCCGAGAGATA 110169

QY 61 CGCAGACCCACCTGTGTATCATCTGGCCCGAGGTGGCAGCATGGCGCAATGAGAAC 120
Db 110168 CGCAGACCCACCTGTGTATCATCTGGCCCGAGGTGGCAGCATGGCGCAATGAGAAC 110109

QY 121 AGCATCAGCAGGCGCCAGGAGGCTTACAGGTGGCTGCGGAGGAGCCAGGACTCACCT 180
Db 110108 AGCATCAGCAGGCGCCAGGAGGCTTACAGGTGGCTGCGGAGGAGCCAGGACTCACCT 110049

QY 181 CAAGCTCTAAACCAATCACACCCCTTACTTACACAGAAAGCTCAGAGTGCCTCAAG 240
Db 110048 CAAGCTCTAAACCAATCACACCCCTTACTTACACAGAAAGCTCAGAGTGCCTCAAG 109989

QY 241 TCGCCACCCGCTGCGAGAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCTGAG 300
Db 109988 TCGCCACCCGCTGCGAGAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCTGAG 109929

QY 301 GTTTCTCACATCAAAAAGAGAGGTGTCCAAAACGGTGTGTGTCAGAGACTTACGAGAGA 360
Db 109928 GTTTCTCACATCAAAAAGAGAGGTGTCCAAAACGGTGTGTGTCAGAGACTTATGAGAGA 109869

QY 361 GGAGGGACGTGAGCCCTCAGCCACAGGTACAGAGGAGTGTGTGTGTGTGAACCT 420
Db 109868 GGAGGGACGTGAGCCCTCAGCCACAGGTACAGAGGAGTGTGTGTGTGTGAACCT 109809

QY 421 GGGCAGCCAGAGATGACATTGACAGATCTCTCCACAGCCACGCTCCCAACCGGAGG 480
Db 109808 GGGCAGCCAGAGATGACATTGACAGATCTCTCCACAGCCACGCTCCCAACCGGAGG 109749

QY 481 AGAAATGTGCAACCTGGTGTGTGAGTAAACCAAGGGCTGGAGTGTGAGCAGGAG 540
Db 109748 AGAAATGTGCAACCTGGTGTGTGAGTAAACCAAGGGCTGGAGTGTGAGCAGGAG 109689

QY 541 GAGCCCATGTGAGGAGTGACAGCTAGACACAGAGGACAGCGCTTGGAGGAGAGGCT 600
Db 109688 GAGCCCATGTGAGGAGTGACAGCTAGACACAGAGGACAGCGCTTGGAGGAGAGGCT 109629

QY 601 GAGGAGGCGCCGAGCAGGATGGAGTGCAGTGTGTGTGTGAGATCAAGCCCTTGTG 660
Db 109628 GAGGAGGCGCCGAGCAGGATGGAGTGCAGTGTGTGTGTGAGATCAAGCCCTTGTG 109569
```

```
QY 661 CCTCCCCAG 669
Db 109568 CCTCCCCAG 109560

RESULT 11
AC138715 164806 bp DNA linear ROD 25-NOV-2003
LOCUS Mus musculus BAC clone RP23-283E23 from chromosome 15, complete
DEFINITION sequence.
AC138715
AC138715.3 GI:38093886
VERSION HTG.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 164806)
AUTHORS Bauer,H., Shah,N., Kozlowski,A., Haglund,K. and Haakenson,W.
TITLE The sequence of Mus musculus BAC clone RP23-283E23
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 164806)
AUTHORS Wilson,R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 164806)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (15-JAN-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 164806)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (25-FEB-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 164806)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 164806)
AUTHORS Wilson,R.
TITLE Direct Submission
JOURNAL Submitted (25-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Oct 31, 2003 this sequence version replaced gi:27764802.
```

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: submissions@watson.wustl.edu
----- Summary Statistics
Center project name: M_BA0283E23

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see

Db 12540 GAGGAGCGCTGAGCAGATGAGCGGCTGTGGCTCTGCCAGGATCAAAAGCCCTTG 12599

OY 661 CCTCCAG 669

Db 12600 CTCTCCAG 12608

RESULT 13

AC099736/c

DEFINITION Mus musculus chromosome 15, clone RP23-228F5, complete sequence.

ACCESSION AC099736

VERSION AC099736.10 GI:30841061

KEYWORDS HTG.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 186137)

AUTHORS Birren, B., Nusbaum, C. and Lander, E.

JOURNAL Mus musculus chromosome 15, clone RP23-228F5

TITLE Unpublished

REFERENCE 2 (bases 1 to 186137)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Bouckhalter, B., Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Riback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

TITLE Submitted (19-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

JOURNAL 3 (bases 1 to 186137)

REFERENCE 3 (bases 1 to 186137)

AUTHORS Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Atachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckhalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArelano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

TITLE Submitted (30-SEP-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

JOURNAL 3 (bases 1 to 186137)

REFERENCE 3 (bases 1 to 186137)

AUTHORS Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Atachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckhalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArelano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

TITLE Submitted (30-SEP-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

JOURNAL 3 (bases 1 to 186137)

REFERENCE 3 (bases 1 to 186137)

AUTHORS Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Atachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckhalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArelano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

TITLE Submitted (01-MAY-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

JOURNAL 4 (bases 1 to 186137)

REFERENCE 4 (bases 1 to 186137)

AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Atachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckhalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArelano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (17-MAY-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

5 (bases 1 to 186137)

AUTHORS Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Atachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckhalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArelano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, X., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (30-SEP-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On May 17, 2003 this sequence version replaced gi:30270708.

All repeats were identified using RepeatMasker:

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Smit, A.F.A. & Green, P. (1996-1997)

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L17149

Center clone name: 228_F_5

FEATURES

Location/Qualifiers

1..186137

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="15"

/map="15"

/clone="RP23-228F5"

/clone_lib="RPCI-23 Female Mouse BAC"

1..6

/note="clone boundary clone end:SP6 end site:ECORI"

misc_feature

repeat_region	757. 825	/rpt_family="GA-rich"	Db	127967	CGCAGCAACCCCTGGTTATCAATTTGGCCCGAGGTTGGCAGCACTGGCGGAATGAGAAC	127908
repeat_region	834. 951	/rpt_family="B1_MW"	Qy	121	AGCATCAGCAGGCCCGAGGAGCTACAGCTGGCTGGCTGGCGGAGGAGCCAGGACTCACCT	180
repeat_region	1043. 1078	/rpt_family="CAAA)n"	Db	127907	AGTACCAACAGGCCCGAGGAGCTGCAGCTGGCTGGCGGAGCAACTCATGACTACT	127848
repeat_region	1222. 1255	/rpt_family="(TA)n"	Qy	181	CAAGCTCTCTAAACCAATCACACCCCTACTTTCACACCAAGAGCTCAGAGTGCCTCCAAAG	240
repeat_region	1256. 1278	/rpt_family="(CA)n"	Db	127847	ACGCTCTCTAAAGAGCGGTCC-----TTACCAGACTGCCCCCAAA	127806
repeat_region	1356. 1463	/rpt_family="(TA)n"	Qy	241	TGCGACCCCGCTGCCAGAGGACATGGAGATGACAAAGCTCAGAGAAAGCCCTGAG	300
repeat_region	1744. 1807	/rpt_family="(TG)n"	Db	127805	ACTCTGTCTCCAAAGCCAGATCGAGACGAGAGGAGCAACACTCAGAAGAGCCACCGAG	127746
repeat_region	1987. 2116	/rpt_family="L2"	Qy	301	GTCTTCTCATCAAAAAGAAAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGA	360
repeat_region	2180. 2214	/rpt_family="AT-rich"	Db	127745	GTCTCCCATTTAAAGGAAAGAGGTGACCAACGGTTGTTCAGCAAGCTTATGAGAGG	127686
repeat_region	2219. 2607	/rpt_family="L2"	Qy	361	GGAGGGAGCTGAGCCACTCAGCCACAGGTACGAGAGGATGCTGGTGTGCTTCAACT	420
repeat_region	complement(3071. 3179)	/rpt_family="L2"	Db	127685	GGAGGAGATGTGAACCTACCTGAGCCACAGGTATGAGAAATGATGGCGGCTGTCTGAAGCT	127626
repeat_region	4246. 4358	/rpt_family="PB1D7"	Qy	421	GGGCAGCCAGAGATGACATTGCACAGAAATCCTCCACAGCCAGCGGTCCCAACGGCGAGG	480
repeat_region	4903. 5073	/rpt_family="B3"	Db	127625	ATTCAGCCAGAGATGACATTGCACAGAAATCCTTCTTAGTCAGACTCGCCACACGAGAG	127566
repeat_region	5076. 5104	/rpt_family="AT-rich"	Qy	481	AGAAATGTGCCAACTGGTGTCTGAGCTAACCAAGGGCTGGAGAGTATGGAGCAGGAG	540
repeat_region	complement(6065. 6297)	/rpt_family="L1MA4A"	Db	127565	AGAAATGCACCAACCTGGTGTCTGAGCTGACAAAGGCTGGAAGTATGATGGAACAGGAA	127506
repeat_region	complement(6380. 6532)	/rpt_family="MIR"	Qy	541	GAGCCCATGAGGAGGTGACAGCTAGACAGAGGAGCAGCGGTATGAGAGGAGGCT	600
repeat_region	7597. 7701	/rpt_family="PB1D10"	Db	127505	GAGCCCATGGAAGAGTACAGCGGTAGACAGAGGACAGTGGCTACGAGGGATATG	127446
repeat_region	7702. 7782	/rpt_family="ID4"	Qy	601	GAGGAGGCGCGAGCAGGATGGAGTGGCTGGTGGTCAGGATCAAGCGCCCTTG	660
repeat_region	12709. 12786	/rpt_family="ID3"	Db	127445	GAGGAGGCGCTGAGCAAGATGACGCGCTGTGGTCTCTGCGAGGATCAAGCGCCCTTG	127386
repeat_region	13119. 13156	/rpt_family="AT-rich"	Qy	661	CCCTCCAG 669	
repeat_region	complement(13894. 15060)	/rpt_family="Lx"	Db	127385	CTCTCCAG 127377	
repeat_region	complement(15419. 15545)	/rpt_family="Lx6"				
repeat_region	complement(15559. 15771)	/rpt_family="RMER17B"				
repeat_region	15623. 16180	/rpt_family="RMER13B"				
repeat_region	complement(17622. 17987)	/rpt_family="ORR1B1"				
repeat_region	complement(18208. 18599)	/rpt_family="Lx4"				
repeat_region	18644. 18971	/rpt_family="Lx5"				
repeat_region	complement(19446. 19622)	/rpt_family="ORR1D"				
repeat_region	complement(19636. 19865)	/rpt_family="Lx4"				
repeat_region	complement(19885. 20149)	/rpt_family="L1_MW"				
repeat_region	complement(20156. 20335)	/rpt_family="ORR1D"				
repeat_region	complement(20788. 20953)					
Query Match	36.6%;	Score 419.8;	DB 10;	Length 186137;		
Best Local Similarity	78.3%;	Pred No.1.8e-104;				
Matches 524;	Conservative 0;	Mismatches 127;	Indels 18;	Gaps 1;		
Qy	1	ATGGCTCCGGCGGAAAGGAAAGCGGGAGGCGCCAGCAAGAGCGCCCTCCGGAAGATA	60			
Db	128027	ATGGCTCCAGGAGAAAGGAAAGGAGGCGGGCGGCGCAAGAGTGCCCTCCGGAAGTGC	127968			
Qy	61	CGCACAGCACCCCTGGTTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGCGGCGATGAGAAC	120			

RESULT 14

AC136853/c

LOCUS

DEFINITION

AC136853

AC136853.1 GI:24818534

VERSION

HTG: HTGS PHASE1

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

1 (bases 1 to 73341)

AUTHORS

Muzny, D.W., Metzker, M., Lee, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D., Araya-Debechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dunn, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Muzny, D.W., Metzker, M., Lee, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D., Araya-Debechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dunn, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Quevara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Haylak, P., Hines, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowalski, C., Kraft, C.I., Lebowitz, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensuhera, L., Louised, H., Lozano, R.J., Lu, X., Ma, J.,
 Mangum, B., Mapua, P., Martin, K., Mahmoud, M., Malloy, K., Mangum, A.,
 Mawhinney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A.,
 Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,
 Morris, S., Mundasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D.,
 Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okunolu, G.,
 Olarunpungu, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,
 Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A.,
 Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quintero, J., Rachlin, E.,
 Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,
 Reuter, M., Richards, S., Riggs, P., Rives, C., Rodkey, T., Rojas, A.,
 Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savory, G., Scherer, S.,
 Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,
 Sisson, I., Sitter, C.D., Smay, D., Sneed, A., Sodergren, E.,
 Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A.,
 Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,
 Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D.,
 Waldron, L., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J.,
 Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wlaczek, R.,
 Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,
 Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,
 Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R.,
 Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
 Unpublished
 2 (bases 1 to 73341)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: KCVW
 Center clone name: CH230-263K14
 ----- Summary Statistics
 Sequencing vector: Plasmid/
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 28037 bases at least Q40
 Consensus quality: 29795 bases at least Q30
 Consensus quality: 31083 bases at least Q20

----- NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently
 consists of 40 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1
 1192: contig of 1192 bp in length
 1292: gap of unknown length
 1293: contig of 1215 bp in length
 2507: gap of unknown length
 2608: contig of 1506 bp in length
 4113: contig of 1506 bp in length
 4213: gap of unknown length

4214
 5289: contig of 1076 bp in length
 5389: gap of unknown length
 5589: contig of 1137 bp in length
 6626: gap of unknown length
 6627: contig of 1134 bp in length
 7860: gap of unknown length
 9123: contig of 1263 bp in length
 9223: gap of unknown length
 10318: contig of 1095 bp in length
 10418: gap of unknown length
 11713: contig of 1295 bp in length
 11813: gap of unknown length
 13025: contig of 1212 bp in length
 13125: gap of unknown length
 14223: contig of 1197 bp in length
 14423: gap of unknown length
 15633: contig of 1217 bp in length
 15733: gap of unknown length
 16850: contig of 1111 bp in length
 16950: gap of unknown length
 18316: contig of 1366 bp in length
 18416: gap of unknown length
 20103: contig of 1687 bp in length
 20203: gap of unknown length
 21733: contig of 1530 bp in length
 21833: gap of unknown length
 23180: contig of 1347 bp in length
 23280: gap of unknown length
 25124: contig of 1844 bp in length
 25224: gap of unknown length
 26789: contig of 1565 bp in length
 26889: gap of unknown length
 28009: contig of 1120 bp in length
 28109: gap of unknown length
 29909: contig of 1800 bp in length
 30009: gap of unknown length
 31356: contig of 1347 bp in length
 31456: gap of unknown length
 32133: contig of 1677 bp in length
 32333: gap of unknown length
 34330: contig of 1197 bp in length
 34530: gap of unknown length
 36398: contig of 1868 bp in length
 36999: gap of unknown length
 37783: contig of 1285 bp in length
 37883: gap of unknown length
 40081: contig of 2198 bp in length
 40181: gap of unknown length
 41870: contig of 1689 bp in length
 41970: gap of unknown length
 44027: contig of 2057 bp in length
 44127: gap of unknown length
 46329: contig of 2202 bp in length
 46429: gap of unknown length
 47928: contig of 1499 bp in length
 48028: gap of unknown length
 50367: contig of 2339 bp in length
 50467: gap of unknown length
 52413: contig of 1946 bp in length
 52513: gap of unknown length
 55446: contig of 2333 bp in length
 55546: gap of unknown length
 57927: contig of 2381 bp in length
 58027: gap of unknown length
 58802: contig of 1775 bp in length
 59302: gap of unknown length
 62792: contig of 2890 bp in length
 62892: gap of unknown length
 66868: contig of 3976 bp in length
 66968: gap of unknown length
 70615: contig of 3647 bp in length
 70715: gap of unknown length
 73341: contig of 2626 bp in length

```
FEATURES
Source
Location/Qualifiers
1..73341
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-263K14"

ORIGIN
Query Match 32.5%; Score 373; DB 2; Length 73341;
Best Local Similarity 85.6%; Pred. No. 1.5e-91;
Matches 415; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 662 CTTCCAGGTAAGACAGATTACAGAGAACTCAACTGCAAGCCCAACAGAAATATATGCC 721
DB 64707 CTTCTAGGCGAAACAGATACACTGCTGTAAGGCCCATCGGAATATACAGCC 64648

QY 722 CAGTGGCAACTTGAAGGGAGATGGCAGAGTGGCTGTAGACACATACATCCACGA 781
DB 64647 AAGTGACAACTTGAAGGGCAGGTGGCAGCAGTGGCGGATGAACACATACAGTCACAGA 64588

QY 782 AGCTCAATCCCTTTCAGTCAAGAGTTTGATTACGAGCTGGCCATGTCCACCGCCCTACACA 841
DB 64587 AGCTCAACCCCTTCAGTGTGATTTGACTATGACCTAGCCATGTCACACTGCCATCCACA 64528

QY 842 AAGGAGATGAGGGCTATGGCCGCCGCCCAAGAGAACCAAACTGCTGAAGGGCCCAAGC 901
DB 64527 AAGGAGATGAGGGCTATGGCCGCCGCCCAAGAGAACCAAACTGCTGAAGGGCCCAAGC 64468

QY 902 GTGCTGAGGAGCAGATCTACAGGAAATGATGGACATGCTTCAATTCGCAATGG 961
DB 64467 GAGCCGAGGAGCAGATCTATCGGAAATGATGGATTTGCTTTATCCGCAATGG 64408

QY 962 CTCGCCAGAGCAGATCGCAAGATCCAGGTTACTTTTGGAGATCTTTTACAGATACG 1021
DB 64407 CTCGCCATAGCAGATCGCAAGATCCAGGTTACTTTTGGAGATCTTTTACAGATACG 64348

QY 1022 TTCGATTTTCAGATTAAGTATGAGGATCTCATCGTCCAGGAAACATGAGTGGTAG 1081
DB 64347 TTCGATTTTCAGATTAAGTATGAGGATCTCATCGTCCAGGAAACATGAGTGGTAG 64288

QY 1082 ACTTTGAAGGAGATGCTATGGCAAGGCCGAGATGACCATGTTGTGATTACGCTACTCA 1141
DB 64287 ACTTTGAAGGAGATGCTATGGCAAGGCCGAGATGACCATGTTGTGATTACGCTACTCA 64228

QY 1142 AGTGA 1146
DB 64227 AGTAA 64223

RESULT 15
AC114181
LOCUS
DEFINITION
Rattus norvegicus clone CH230-19304, WORKING DRAFT SEQUENCE, 2
unordered pieces.
AC114181
VERSION
AC114181.5 GI:30579540
KEYWORDS
HTG: HTGS_PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 220702)
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebech, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy, Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Georgiev, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowals, C., Kratt, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenschew, L., Loulseghe, H., Lozano, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montenayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwankwelu, O., Okwuonu, G., Olarnpungoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reid, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,
Snead, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlaczek, R., Woodson, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 220702)
Worley, K. C.
Direct Submission
Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 220702)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:24813086.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly ('a' contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
-----
Center project name: GLPO
```

Center clone name: CH230-19304

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 214139 bases at least Q40

Consensus quality: 215944 bases at least Q30

Consensus quality: 217301 bases at least Q20

Estimated insert size: 225948; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 2 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 219530: contig of 219530 bp in length

* 219531 219630: gap of unknown length

* 219631 220702: contig of 1072 bp in length.

----- Location/Qualifiers

1. .220702

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-19304"

1. .1179

/note="wgs_end_extension"

clone_end:Sp6"

complement(1697..2319)

/note="clone boundary"

clone_end:Sp6"

site:EcoRI

end sequence:BH314401"

complement(218360..218637)

/note="clone boundary"

clone_end:T7

site:EcoRI

end_sequence:BH314400"

ORIGIN

Query Match 32.5%; Score 373; DB 2; Length 220702;

Best Local Similarity 85.8%; Freq. No. 1.5e-91;

Matches 415; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 662 CTCTCCAGGTAAACAGATTACAGAGAACTCAACTGCAAGCCCAACAGAAATATAGCC 721

Db 213689 CTTCTAGGGCAACACAGATCTCAGAGACACTCAACTGTAAGGCCCATCGAAATACAGCC 213748

QY 722 CAGTGGCACTTGAAGGAGATGGCAGAGTGGGCTGATGAACATACATCCAGA 781

Db 213749 AAGTGGACAACITTGAAGGAGGTGGCAGCAGTGGGCCGATGAACATACAGTCACAGA 213808

QY 782 AGCTCAATCCTTTTCAGTGAAGAGTTGATTACAGAGTGGCCATGTCCACCCGCTACACA 841

Db 213809 AGCTCAACCCCTTCAGTGTGATTTGACTATGACCTAGCCTAGTCCACTCGACTCCACA 213868

QY 842 AAGGAGATGAGGGCTATGGCCGCCCCCAAGAGAACCAAACTGCTGAAGGGCCCAAGC 901

Db 213869 AGGGAGACGAGGGATATGGCCGCCCAAGAGAGGGAAGCAAGACAGCTGAAGGGCCCAAGA 213928

QY 902 GTGCTGAGGAGCATCTCAGGGAATGATGGACATGTCTTCATTCTGCACAAATGG 961

Db 213929 GAGCCGAGGAGCATTCTATCGGGAATATTGGAATGTGCTTTGTTATCCGCACATGG 213988

QY 962 CTCGCCACAGACGAGATGGCAAGATCCAGGTACTTTTGGAGATCTCTTGAAGATACG 1021

Db 213989 CTCGCCATAGACGAGATGGCAAGATCCAGGTACTTTTGGAGATCTCTTGAAGATACG 214048

QY 1022 TTCGTATTTCAGATAAAGTAGTGGGCATTTTCATGCGTGCAGGAACATGGACTGTAG 1081

Db 214048

Db 214049 TTCGAATTTTCAGATAAAGTGGTGGGATCCTCATGCGTCCAGGAACACGAGCTGGTGC 214108

QY 1082 ACTTTGAAGGAGAGATGCTATGGCAAGGGCGGAGATGACCATGTTGTGATTAGCTACTCA 1141

Db 214109 ACTTTGAAGGGGAGATGCTATGSCAAGGCAAAAGATGACCATGTTGTGATTACTCTCCTTG 214168

QY 1142 AGTGA 1146

Db 214169 AGTAA 214173

Search completed: May 5, 2004, 02:13:46

Job time : 4467 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 4, 2004, 08:49:08 ; Search time 64 Seconds
(without alignments)
11299.496 Million cell updates/sec

Title: US-10-644-659A-1

Perfect score: 2058
Sequence: 1 agggctccggcgcaaaagga.....tgattacgctactcaagtga 1146

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cgn2_1/USPTO_spool.p/US10644659/runat_04052004_084902_28522/app_query.fasta_1.1287
-DB=SPTRMBL_25 -OPWT=Eastan -SUPFI=n2p -rpt -MINMATCH=0_1 -LOOPL=0
-LOOPEXT=0 -FNIS=bits -SPART=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi
-LIST=45 -LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTWRT=pt -NORMEXT -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10644659 @CGN_1_1_86 @runat_04052004_084902_28522 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTRMBL_25:

1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phage:
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_rvirus:
16: sp_bacteria:
17: sp_archaea:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2029	98.6	381	4 Q8N0Z2	Q8N0Z2 homo sapien

ID	Q8N0Z2	PRELIMINARY;	PRT;	381 AA.
AC	Q8N0Z2;			
DT	01-OCT-2002 (TrEMBLrel. 22, Created)			
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)			
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)			
DE	STARS (Hypothetical protein FLJ35375).			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=22086201; PubMed=11983702;			
RA	Arai A., Spencer J.A., Olson E.N.;			
RT	"STARS, a Striated Muscle Activator of Rho Signaling and Serum Response Factor-dependent Transcription.";			
RL	J. Biol. Chem. 277:24453-24459(2002).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Skeletal muscle;			
RA	Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,			
RA	Fukumami Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,			
RA	Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,			

ALIGNMENTS

RESULT 1
Q8N0Z2
ID Q8N0Z2
AC Q8N0Z2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE STARS (Hypothetical protein FLJ35375).

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22086201; PubMed=11983702;

RA Arai A., Spencer J.A., Olson E.N.;

RT "STARS, a Striated Muscle Activator of Rho Signaling and Serum Response Factor-dependent Transcription.";

RL J. Biol. Chem. 277:24453-24459(2002).

RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukumami Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,

Db 355 HisPheGluGlyGluMetLeuTrpGlnGlyAspHisValValIleThrLeuLeu 374
 QY 1141 AAG 1143
 Db 375 Glu 375

RESULT 5
 Q8BLH3 PRELIMINARY; PRT; 335 AA.

ID	Q8BLH3;
AC	01-WAR-2003 (TReMBLrel. 23, Created)
DT	01-WAR-2003 (TReMBLrel. 23, Last sequence update)
DT	01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE	Weakly similar to F5672.1 protein (Fragment).
GN	C130068012Bik.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC	Mammalia; Eutheria; Rodentia; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
[1]	SEQUENCE FROM N.A.
RN	STRAIN=C57BL/6J;
RC	MEDLINE=22354683; PubMed=12466851;
RX	The FANTOM Consortium,
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;
RA	"Analysis of the mouse transcriptome based on functional annotation of
RT	60,770 full-length cDNAs.";
RL	Nature 420:563-573 (2002).";
DR	EMBL; AK045186; BAC32252.1; -;
DR	MGI; MGI:2444891; C130068012Rik.
FT	NON TER 1
SQ	SEQUENCE 335 AA; 38372 MW; EBI5181FAB749212 CRC64;

Alignment Scores:
 Pred. No.: 2,93e-106 Length: 335
 Score: 1415.00 Matches: 267
 Percent Similarity: 85.92% Conservative: 26
 Best Local Similarity: 78.30% Mismatches: 42
 Query Match: 68.76% Indels: 6
 DB: 11 Gaps: 1

US-10-644-659A-1 (1-1146) x Q8BLH3 (1-335)

Qy	121 AGCATCAGCAGCGCCCGAGAGCTACAGCTGGTGCGGGAGGCCAGACTCACT 180
Dd	:::::
Qy	1 SerThrLysslnAlagIngluProalaglylrPeuProglyAlaThrHisAspLeuPro 20
Qy	181 CAAGCTTCCTAAACCATCACACCCCCTACTTCACACCAGAAGGTCTAGAGTCCCCAAAG 240
Dd	:::::
Dd	21 AsnAlaProLysGluAlaGlyPro-----TyrGlnHisAlaProLys 34
Qy	241 TCGCCACCCCGCTGCACAGGACATGGAGATGGACAAGCTCAGAAAAGCCCTGAG 300
Dd	:::::
Dd	35 ThrLeuSerProLysProAspargAspglyGluGlyGlnHisSerGluGluAlaThrGlu 54
Qy	301 GTTTCTCATTCAAAGAAGAGGTGTCCAAAACGGTGGTCCACAGACTTACGAGAGA 360
Dd	:::::
Dd	55 ValSerHistlelyslargylsGluValThrArgThrValValSerLysAlaTyrlguArq 74
Qy	361 GGAGGGAGCTGAGCCACCTCAGGCACAGGTACGAGGGATGTGTGTGCTTGAACCT 420
Dd	:::::
Dd	75 GlyGlyAspValasnTyrlserHisargTyrlGluasnAspglyGlyValSerGluAla 94
Qy	421 GGGCAGCCAGAGAAATGACATTGACAGAAATCTTCACAGCCACGGCTCCCAACGCGAGG 480
Dd	:::::
Dd	95 IleGlnProgluasnAspileaspArglleLeuLeuSerHisAspSerProThrArqarg 114
Qy	481 AGAAATGTCCCAACTGGTGTCTGAGCTTAACCAAGGGCTGGAGAGTGTGAGAGAGGCT 540
Dd	:::::
Dd	115 ArglylsCysThrAsnLeuvalSerGluLeuThrlyslsGlylTryllysValMetGluGlnGl 134
Qy	541 GAGCCCATGCGAGGATGACAGCGTAGACACAGAGGACACGGCGTGTATGAGGAGAGGCT 600

Alignment Scores: 3.68e-14 Length: 371
pred. No.:

Thu May 6 09:39:12 2004

Best Local Similarity: 39.16% Mismatches: 49
Query Match: 11.10% Indels: 11
DB: 5 Gaps: 4

US-10-644-659A-1 (1-1146) x Q22133 (1-289)

QY 721 CCAGTGGGCAACTTTAAAGG-----AGATGCACGAGTGCGCTGTATGAACAATA 771
|||::|::|::|::|::|::|::|::|::|::|
Db 2 ProIleGlySerAlaSerAspThrIleArgLysPheAsnAlaValAlaGlnAlaAsnGlu 21

QY 772 CAATCCAGAAGCTCAATCTTCAGTGAAGACTTTTCATTACAGCTGCCCATCACC 831
::|::|::|::|::|::|::|::|::|::|
Db 22 GluValLeuLysAsnProTyrsrAspThrytyrLysileGln---AlaPreAspThr 40

QY 832 CGCTACACAAAGAGATGAGGCGCTATGCCGCCCCCRAAGAAAGAACCAAACCTGCGAA 891
::|:
Db 41 Lys-----AsnTyrgLyArgProProGlyserLysThyrGluAla 54

QY 892 AGGCCAAGCGTCTGAGGAGCACATCTACAGGGAATGATGACATGTGCTTCATTATC 951
|||::|::|::|::|::|::|::|::|::|::|
Db 55 ArgGlyleLysalaGlyValHisValcysArgGluIlellePheLeuLysGluThrlle 74

QY 952 TGCACAAATGCTGCCACACACAGATGCCAAGATCCAGGTACTTTTTGGAGATCTCTTT 1011
::|:
Db 75 AspSerAsnAlaaspGlyGluProHisLys--TyrVallysPheGlyLysLeuPhe 93

QY 1012 GACAGATACGCTTGTATTTTTCAGATAAAGTAGTGGGCATTCTCATCGCTGCCAGGAACAT 1071
::|:
Db 94 AsniletyrSerPhetyrSerAspLysLeuValglyMetLeulleAr-galaaargLysTy 113

QY 1072 GGACTGCTAGACTTTGAAGAGAGATGCTATGGCAAGCGCCAGATGACCATGTGTGTGATT 1131
::|:
Db 114 GlyLeuValHisPheGluGlyGluMetLeuTyrgLnarginAspAspGlulylsilile 133

QY 1132 ACGCTACTC 1140
|||::|::|::|
Db 134 ThrMetLeu 136

RESULT 9

Q9VZW6 PRELIMINARY; PRT; 153 AA.

ID AC Q9VZW6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Cg2113 protein (AT18037p).
GN Cg2113.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyridae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Cocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Gallie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Workman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Beran B.P., Bhadrari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Crawley S., Dahlike C., Davenport L.B., Davies P.,
RA de Fabios B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garq N.S., Gelbart W.M., Glasser K.,
RA Glodsek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

Score:	279.50	Matches:	64
Percent Similarity:	61.81%	Conservative:	25
Best Local Similarity:	44.44%	Mismatches:	50
Query Match:	13.58%	Indels:	5
DB:	5	Gaps:	2

US-10-644-658A-1 (1-1146) x O46052 (1-371)

QY	718	AGCCACAGTGGGCAACTTGAAGGGAGATGGCAGCAGTGGCGTGATGAACACATACATCC	777
		::: :::	:::
Db	17	SerProLeuSerSerLysValAlaMetPheAsnGlnAlaThrGlnHisLysGlnSer	36
		::: ::: ::: :::	
QY	778	CAGAAGCTCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTGGCAGTGTCCACCCGGCTA	837
		::: ::: ::: :::	:::
Db	37	GlnLeuLeuAsnProPheSerGlnAsp-----GlyArgAlaAlaSerProLysPro	53
		::: ::: ::: :::	:::
QY	838	CACAAAGGAGATGAGGGCTATGCGCGCCCAAGAGGAACAAAACTGCTGGAAGGCC	897
		::: ::: ::: ::: :::	:::
Db	54	ThrPheSerLysAspGlnTyrGlyLysProLeuAlaGlySerLeuThrGluMetArgGly	73
		::: ::: ::: :::	:::
QY	898	AACGGTGTGAGAGCAGACATCTACAGGGAATGATGGACATGTGCTTCATTATCTGCACA	957
		::: ::: ::: :::	:::
Db	74	GlnLysAlaAsnIleHisValMetLysGluMetLeuGluLeuLysGlnIleIleAsnSer	93
		::: ::: ::: :::	:::
QY	958	ATGGCTCGCCACAGACAGATGGCAAGATCCAG-----GTTACTTTTGGAGATCTCTTT	1011
		::: ::: ::: :::	:::
Db	94	GluGlyTyrAspValLysAspGluProThrMetArgValIleProPheGlyGluLeuPhe	113
		::: ::: ::: :::	:::
QY	1012	GACAGATACCTTCGTTATTCAGATAAGTAGTGGCATTCTCATCGTGGCCAGGAACAT	1071
		::: ::: ::: :::	:::
Db	114	AsnIleTyrAsnTyrIleSerAspLysValValGlyIleLeuLeuArgAlaArgLysHis	133
		::: ::: ::: :::	:::
QY	1072	GGACTGTGTAGACTTTGAAGGAGAGATGCTATGCAAGCGCAGATGACCATGTTGTGATT	1131
		::: ::: ::: :::	:::
Db	134	LysLeuValAspPheGluGlyGluMetLeuTyrGlnArgArgAspAspValProVal	153
		::: ::: ::: :::	:::
QY	1132	ACGCTACTCAAG 1143	
		::: ::: ::: :::	:::
Db	154	PheLeuLeuLys 157	
		::: ::: ::: :::	:::

RESULT 8

Q22133 PRELIMINARY; PRT; 289 AA.

ID	Q22133	AC	Q22133
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DT	01-JUN-2003	(TREMBLrel. 24, Last annotation update)	
DE	T0428.4	protein.	
GN	T0428.4		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]	SEQUENCE FROM N.A.	
RP	Palmer S.;		
RA	Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.		
RL	[2]		
RN	SEQUENCE FROM N.A.		
RP	MEDLINE=99069613; PubMed=9851916;		
RX	RA none;		
RA	"Genome sequence of the nematode C.elegans: A platform for		
RT	investigating biology."		
RL	Science 282:2012-2018(1998).		
DR	EMBL; Z35663; CAA94722.1; -.		
DR	PIR; T24421; T24421.		
DR	WormPep; T04A8.4; CE01073.		
SQ	SEQUENCE 289 AA; 32972 MW; 20DB67F731F03098 CRC64;		

Alignment Scores:

Pred. No.:	4,83e-10	Length:	289
Score:	228.50	Matches:	56
Percent Similarity:	58.04%	Conservative:	27

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,	Score: 192.00	Matches: 44
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,	Percent Similarity: 56.20%	Conservative: 24
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	Best Local Similarity: 36.36%	Mismatches: 49
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,	Query Match: 9.33%	Indels: 4
RA Laakso P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,	DB: 5	Gaps: 2
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,	US-10-644-659A-1 (1-1146) x Q9VZW6 (1-153)	
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,		
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,		
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,		
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,		
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,		
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,		
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,		
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,		
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,		
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,		
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,		
RA Zheng X.H., Zheng F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,		
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,		
RT "The genome sequence of Drosophila melanogaster";		
RL Science 287:2185-2195 (2000).		
RN [2]		
RP SEQUENCE FROM N.A.		
RA Celisner S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,		
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,		
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,		
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,		
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,		
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,		
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,		
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,		
RA Mcintosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,		
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,		
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,		
RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,		
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,		
RT "Sequencing of Drosophila melanogaster genome";		
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.		
RN [3]		
RP SEQUENCE FROM N.A.		
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,		
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,		
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celisner S.E.,		
RA Clamp M., Drysdale R., Emmert D., Frise E., de Gray A., Harris N.,		
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,		
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,		
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,		
RT "Annotation of Drosophila melanogaster genome";		
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.		
RN [4]		
RP SEQUENCE FROM N.A.		
RA Adams M.D., Celisner S.E., Gibbs R.A., Rubin G.M., Venter C.J.,		
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.		
RN [5]		
RP SEQUENCE FROM N.A.		
RA FlyBase:		
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.		
RN [6]		
RP SEQUENCE FROM N.A.		
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,		
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,		
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,		
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,		
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,		
RA Celisner S.,		
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.		
DR EMBL; A5003476; AAF47699.2; -		
DR EMBL; AY113283; AAM29288.1; -		
DR FlyBase; FBgn0035384; CG21113.		
SQ SEQUENCE 153 AA; 17529 MW; F47F7E126DE8DF0 CRC64;		
Alignment Scores:	3.95e-07	Length: 153
Pred. No.:		

QY 784 CTCATCTCTTCAGTGAAGATTTGATTACGAGCTGCCATGTCCACCGCTACACAA 843	Score: 192.00	Matches: 44
DB 3 IleAsnProPheAlaHis-----TyAsnValAsnGluMetProLysArgThrPhe 19	Percent Similarity: 56.20%	Conservative: 24
QY 844 GGAGATGAGGCTATGGCGCCCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 903	Best Local Similarity: 36.36%	Mismatches: 49
DB 20 ProGluGluGluTyArgAlaProThrGlySerLeuSerGluGlnArgSerLeuGln 39	Query Match: 9.33%	Indels: 4
QY 904 GTGAGGAGCAGCATCTACAGGAATGATGACATGTGCTTCATATCTGCACAATGGCT 963	DB: 5	Gaps: 2
DB 40 AlaAsnValaAlaLeuGluGluLeuGlnLeuGlnLeuGlnLeuGlnLeuGlnLeu 59	US-10-644-659A-1 (1-1146) x Q9VZW6 (1-153)	
QY 964 CGCCACAGACGA---GATGGCAAGATCCAGGTTACTTTTGGAGATCTTTTGCAGATAC 1020		
DB 60 ArgAspAspProIleAspGlyArgLysValLeuAlaPheGlyGlnLeuPheGluThrTy 79		
QY 1021 GTTCGTATTCAGATAAAGTAGTGGCATTCTCATCGTCCAGGAACATGGATGGTA 1080		
DB 80 AsnAsnIleSerAspLysLeuLeuAlaThrLeuLeuGlyAlaArgLysTyGlyPheVal 99		
QY 1081 GACTTTTGAAGGAGATGCTATGGCAAGCGGAGATGACCATGTTGTGATTACCTACTC 1140		
DB 100 AspPheSerGlyGluThrLeuPheGlnGlyArgAspThrGluProValArgLeuLeu 119		
QY 1141 AAG 1143		
DB 120 Arg 120		
RESULT 10		
ID 081881 PRELIMINARY; PRT; 161 AA.		
AC 081881		
DT 01-NOV-1998 (TREMBLrel. 08, Created)		
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)		
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DE Hypothetical protein.		
GN T16L1.130 OR F4G33640.		
OS Arabidopsis thaliana (Mouse-ear cress).		
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;		
OC euroids II; Brassicales; Brassicaceae; Arabidopsid.		
OX NCBI_TaxID=3702;		
RN [1]		
RP SEQUENCE FROM N.A.		
RA Obermaier B., Deutschenbaur S., Piravandi E., Hoheisel J., Jesse T.,		
RA Heijnen L., Vos P., Mewes H.W., Mayer K.F.X., Schueller C., Bevan M.,		
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.		
RN [2]		
RP SEQUENCE FROM N.A.		
RA Obermaier B., Deutschenbaur S., Piravandi E., Mewes H.W., Lemcke K.,		
RA Mayer K.F.X.;		
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.		
RN [3]		
RP SEQUENCE FROM N.A.		
RA EU Arabidopsis sequencing project;		
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.		
DR EMBL; AL031394; CAA20577.1; -		
DR PIR; T04981; T04981.		
DR Hypothetical protein.		
SQ SEQUENCE 161 AA; 18337 MW; 4F42A3D064DE8B5 CRC64;		
Alignment Scores:	0.0221	Length: 161
Pred. No.:	133.50	Matches: 35
Score:		

Best Local Similarity:	25.91%	Mismatches:	88
Query Match:	6.41%	Indels:	86
DB:	4	Gaps:	14

US-10-644-659A-1 (1-1146) x Q9UQ39 (1-956)

QY	6	TCGGGGCGAAGAAAGAAACGGGGAGGCC	-----AGCCAGAG	44
Db	265	SerGlyArgArgGlyGluGlyAspAlaProPheSerGluProGlyThrThrSerThrGln	284	
QY	45	CGCCCTC-----CCGGAAGATACGCACAGCCACCGCTGTCATCAGCTTGCCCGAGGTG	98	
Db	285	ArgProSerSerProGluThrAlaThrLysGlnProSerSerProTyrGluAspLysAsp	304	
QY	99	GCAGCAGTGGCGCAATGAGACAGCATCAGCAGGCCCGCAGGACCTACAGGTGGCTGCC	158	
Db	305	LysAspLysLysGluLysSerAlaThrArgProSerProGluLysSerSerThr	324	
QY	159	GGGAGGACCCA-----	GAATCTACCTCA	182
Db	325	GlyProGluProProAlaProThrProLeuLeuAlaGluArgHisGlyGlySerProgl	344	
QY	183	A-----GCTCTTAAACCAATCACACCCCTTACT-----TCACA	215	
Db	344	nProLeuAlaThrThrProLeuSerGlnGluProValAsnProProSerGluAlaSerPr	364	
QY	216	CCAGAAAGCTCAGAGTGGCCCAAGTGCACCCCGCTGCCAGAGGACATCGAGATGG	275	
Db	364	oThrArgAspArgSerProProLysSerProGluLysLeuProGlnSer---SerSerSe	383	
QY	276	ACAAAGCTCAGAGAAAGCCCTGAGGTTCTCATCATCAAAAAG-----	318	
Db	383	rGluSerSerProProSerProGlnProThrLysValSerArgHisAlaSerSerSerPr	403	
QY	319	-----AAAGAGTGTCCAAAACGGTGT	341	
Db	403	oGluSerProLysProAlaProAlaProGlySerHisArgGluIleSerSerSerProth	423	
QY	342	CAGCAAGACTTACGAGAGGA-----GGGAGCTGAGCCACCTCAGCCACAGGTA	392	
Db	423	rSerLysAsnArgSerHisGlyArgAlaLysArgAspLysSerHis-SerHisThrProS	443	
QY	393	CGAGAGGATGCTGTGTGCTTGAACCTGGGCGCCAGAGNATCATTGACA-----	445	
Db	443	erArgArgMet-----GlyArgSerSerProAlaThrAlaLysA	457	
QY	446	-----GAATCTCTCCACGACCGCTCCCAACGGGA-----GGAGAAAATGTGCCAA	494	
Db	457	rglyArgSerArgSerArgThrProThrLysArgGlyHisSerArgSerArgSerProG	477	
QY	495	CCTGGTCTTGAGCTTAACCAAGGCTGGAGAGTGTATGGAGCAGG---AGAGGCCCATG	551	
Db	477	lnrTpArgArgSer-----ArgSerAlaGlnArgTyrGlyArgSerArgSerProGln-	494	
QY	552	GAGGAGTGACGGTAGACACAGCAGGACGCGCTATGGAGGAGGCTGAGGAGGCC	611	
Db	495	-----ArgArgGlyA	498	
QY	612	CGAGCAGGATGGAGTGCAGGTGGCTGTGTCAGGATCA	649	
Db	498	rArgSerProGlnArgArgProGlyTyrSerArgSer	510	

RESULT 14

ID	Q90U40	PRELIMINARY;	PRT; 1262 AA.
AC	Q90U40;		
DT	01-MAY-2000 (TEMBLrel. 13, Created)		
DT	01-MAY-2000 (TEMBLrel. 13, Last sequence update)		
DT	01-OCT-2003 (TEMBLrel. 25, Last annotation update)		
DE	RNA binding protein (Fragment) .		
OS	Homo sapiens (Human) .		
OC	Eukaryota; Metazoa; Chordata;		
OC	Mammalia; Eutheria; Primates;		
OC	Carnivora; Canidae; Canis;		
OC	Canis lupus familiaris;		
OC	Canis lupus familiaris; Euteleostomi;		
OC	Canis lupus familiaris; Hominidae; Homo.		

NCBI_TaxID=9606; [L]
RN SEQUENCE FROM N.A.
RP Ontaki S., Umeki K., Sawada Y.;
RA "Homo sapiens mRNA for RNA binding protein, partial cds.";
RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL: AB016087; BAA83713.1; -.
DR EMBL: AB016087; BAA83713.1; -.
FT NON_TER 1 1262 1262
FT NON_TER 1262 1262
FT SEQUENCE 1262 AA; 138003 MW; 75E76BD73FA14295 CRC64;
SQ

Alignment Scores:
Pred. No.: 0.0416 Length: 1262
Score: 132.00 Matches: 71
Percent Similarity: 36.50% Conservatives: 29
Best Local Similarity: 25.91% Mismatches: 88
Query Match: 6.41% Indels: 86
DB: 4 Gaps: 14

US-10-644-659A-1 (1-1146) x Q9UQ40 (1-1262)

QY 6 TCCGGGCGAAAGAGGAGGGAGGGCC-----AGCCAAGAG 44
DB 90 SerGlyArgArgGlyGluArgAlaProPheSerGluProGlyThrThrSerThrGln 109
QY 45 CGCCCT-----CGGAAGATACGCCAGCCACCCTGGTCATCAGCTTGGCCCGAGTTG 98
DB 110 ArgProSerProGluThrAlaThrLysGlnProSerSerProTyrGluAspLysAsp 129
QY 99 GCACGAGTGGCGAATCAGAACACATCAGGCAGGCCAGAGCCTACAGGCTGGCTGCC 158
DB 130 LysAspLysLysGluLysSerAlaThrArgProSerProSerProGluAspSerThr 149
QY 159 GGGAGGAGCCCA-----GGTCTTAACCAATCACACCCCTACT-----TCACA 215
DB 150 GlyProGluProProAlaProThrProLeuLeuAlaGluArgHisGlyGly-SerProGl 169
QY 183 A-----GCTCTTAACCAATCACACCCCTACT-----TCACA 215
DB 169 nProLeuAlaThrThrProLeuSerGluGluProValasnProSerGluAlaSerPr 189
QY 216 CCAGAAAGCTCAGAGTCCGCCAAAGTCGCCACCCCGCTGCCAAGAGACATGAGATGG 275
DB 189 oThrArgAspArgSerProProLysSerProGluLysLeuProGlnSer---SerSerSe 208
QY 276 ACAAGCTCAGAGAAGCCCTGAGGTTCTCACATCAAAAG----- 318
DB 208 rGluSerSerProProSerProGlnProThrLysValSerArgHisAlaSerSerSerPr 228
QY 319 -----AAAGAGGTGTCCAAACGGTGGT 341
DB 228 oGluSerProLysProAlaProAlaProGlySerHisArgGluLeuSerSerSerProTh 248
QY 342 CAGCAAGACTTACAGAGAGGA-----GGGACGTGAGCCACCTCAGCCACAGGTA 392
DB 248 rSerLysAsnArgSerHisGlyArgAlaLysArgAspLysSerHis-SerHisThrProS 268
QY 393 CGAGAGGATGTGGTGTGCTTGAACCTGGGCGAGCCAGAGATGATTCATGACA----- 445
DB 268 eArgArgMet-----GlyArgSerArgSerProAlaThrAlaLysA 282
QY 446 -----GANTCTCCACACCCAGCGGTCCCAACGGGA-----CGAGAAATGTGCCAA 494
DB 282 tgGlyArgSerArgSerArgThrProThrLysArgGlyHisSerArgSerArgSerProG 302
QY 495 CCTGGTGTCTGAGCTACCAAGGGCTGGAGAGTGTGAGCAGG---AGGAGCCCACATG 551
DB 302 lnThrArgArgSer-----ArgSerAlaGlnArgTrpGlyArgSerArgSerProGln- 319
QY 552 GAGGAGTGACCGCTAGACACAGAGGACAGCGGCTATGGAGGAGAGGCTTGAGAGAGGCC 611
DB 320 -----ArgArgGlyA 323

QY 612 CGAGCAGGATGGAGTGCAGGTGGTGTGGTCAGGATCA 649
|||||
Db 323 rGserArgSerProGlnArgProGlyTrpSerArgSer 335
|||||
RESULT 15
Q9UQ35 PRELIMINARY; PRT; 2752 AA.
ID Q9UQ35;
AC Q9UQ35;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE RNA binding protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtaki S., Umeki K., Savada Y.;
RT "Homo sapiens mRNA for RNA binding protein, complete cds.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016092; BAA83718.1;
DR Genew; HGNC:16639; SRM2.
SQ SEQUENCE 2752 AA; 299672 MW; 109C64F181097123 CRC64;

Alignment Scores:

Pred. No.: 0.0475 Length: 2752
Score: 132.00 Matches: 71
Percent Similarity: 36.50% Conservative: 29
Best Local Similarity: 25.91% Mismatches: 88
Query Match: 6.41% Indels: 86
DB: 4 Gaps: 14

US-10-644-659A-1 (1-1146) x Q9UQ35 (1-2752)

QY 6 TCCGGCGGAAAGGAAAGCGGGAGGCGCC-----AGCCAAGAG 44
|||||
Db 300 SerGlyArgArgGlyGluGlyAspAlaProPheSerGluProGlyThrThrSerThrln 319
|||||
QY 45 CGCCCT-----CCGGAGATGACGACAGCCACCTGGTGCATCAGCTGGCCCGAGGTG 98
|||||
Db 320 ArgProSerSerProGluThrAlaThrLysGlnProSerSerProTyrgluAspLysasp 339
|||||
QY 99 GCAGCAGTGGCGGAATGAAACAGCATCAGGAGCCAGGAGCCATCAGGCTGGCTGCC 158
|||||
Db 340 LysAspLysGlyLysSerAlaThrArgProSerProSerProGluArgSerThr 359
|||||
QY 159 GGGAGGACCCA-----GGCTCACCTCA 182
|||||
Db 360 GlyProGluProProAlaProThrProLeuLeuAlaGluArgHisGlyGly-SerProgl 379
|||||
QY 183 A-----GCTCTAAACCAATCACACCCCTACT-----TCACA 215
|||||
Db 379 nProLeuAlaThrThrProLeuSerGlnGluProValAsnProProSerGluAlaSerPr 399
|||||
QY 216 CCGAAAGTCAGATGCCCAAGTCGCGACCCCGCTGCCAGAGGACATGGAGATGG 275
|||||
Db 399 oThrArgAspArgSerProProLysSerProGluLysLeuProGlnSer---SerSerSe 418
|||||
QY 276 ACAAGCTCAGAGAAAGCCCTGAGGTTTCTCATCATCAAAAAG----- 318
|||||
Db 418 rGluSerSerProProSerProGlnProThrLysValSerArgHisAlaSerSerSerPr 438
|||||
QY 319 -----AAAGGTGTCCAAAACCGTGGT 341
|||||
Db 438 oGluSerProLysProAlaProAlaProGlySerHisArgGluLeuSerSerSerProth 458
|||||
QY 342 CAGCAAGACTTACGAGAGGA-----GGGACGTGAGCCACCTCAGCCACAGGTA 392
|||||
Db 458 rSerLysAsnArgSerHisGlyArgAlaLysArgAspLysSerHis-SerHisThrProS 478
|||||
QY 393 CGAGAGGGATGCTGGTGTGTTGAACCTGGCGACCCAGCAATGACATTGACA----- 445
|||||

Db 478 erArgArgMet-----GlyArgSerArgSerProAlaThrAlaLysA 492
|||||
QY 446 -----GAATCTCTCCACAGCCACGGCTCCCCCAACGCGGA-----GGAGAAATGTGCCAA 494
|||||
Db 492 xgGlyArgSerArgSerArgThrProThrLysArgGlyHisSerArgSerArgSerProG 512
|||||
QY 495 CCTGGTGTCTGAGCTAAACCAAGGGCTGGAGAGTGTGGAGCAGG---AGGAGCCACATG 551
|||||
Db 512 lnTrpArgArgSer-----ArgSerAlaGlnArgTrpGlyArgSerArgSerProGln- 529
|||||
QY 552 GAGGAGTGCAGCGTAGACACAGAGGACAGCGGCTATGGAGGAGAGGCTTGAGAGAGGCC 611
|||||
Db 530 -----A-rg-rgGlyA 533
|||||
QY 612 CGAGCAGGATGAGTGCAGGTGGCTGTGGTCTCAGGATCA 649
|||||
Db 533 rGserArgSerProGlnArgProGlyTrpSerArgSer 545
|||||

Search completed: May 4, 2004, 08:54:52

Job time : 76 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2004, 23:55:43 ; Search time 2861 Seconds
(without alignments)
11961.572 Million cell updates/sec

Title: US-10-644-659A-1

Perfect score: 1146

Sequence: 1 atggctccggcgaaagga.....tgattacgctactcaagtga 1146

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:**

2: em_esthum:**

3: em_estlin:**

4: em_estov:**

5: em_estov:**

6: em_estpl:**

7: em_estro:**

8: em_estro:**

9: gb_est1:**

10: gb_est2:**

11: gb_est3:**

12: gb_est4:**

13: gb_est5:**

14: gb_est6:**

15: em_estfun:**

16: em_estom:**

17: em_gss_hum:**

18: em_gss_inv:**

19: em_gss_pin:**

20: em_gss_vrt:**

21: em_gss_fun:**

22: em_gss_lam:**

23: em_gss_mus:**

24: em_gss_pro:**

25: em_gss_rod:**

26: em_gss_phg:**

27: em_gss_vrt:**

28: gb_gss1:**

29: gb_gss2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	792.8	69.2	1763	11 AK081696	AK081696 Mus muscu
2	696.8	60.8	2149	11 AK045186	AK045186 Mus muscu
3	618	53.9	618	29 AY399108	AY399108 Homo sapi
4	616.4	53.8	618	29 AY399109	AY399109 Pan trogl

5	501.2	43.7	896	10 BF670302	BF670302 602149566
6	494.8	43.2	717	13 BX671002	BX671002 BX671002
7	481.4	42.0	537	13 BX500732	BX500732 DKZP779K
8	474	41.4	618	29 AY399110	AY399110 Mus muscu
9	429	37.4	576	9 AI605651	AI605651 ma51d03.Y
10	424.6	37.1	909	13 BX667447	BX667447 BX667447
11	423.8	37.0	684	13 BX730420	BX730420 BX730420
12	418.2	36.5	624	13 BX675669	BX675669 BX675669
13	416.6	36.4	553	10 BF549525	BF549525 UI-R-C1-1
14	401	35.0	630	13 BX671003	BX671003 BX671003
15	392.6	34.3	528	13 BX511434	BX511434 BX511434
16	379.6	33.1	582	14 CF368915	CF368915 853657.MA
17	377.8	33.0	646	10 BB643410	BB643410 BB643410
18	363.2	31.7	532	14 CF367562	CF367562 842321.MA
19	358.8	31.3	623	9 AI510175	AI510175 ml97a10.Y
20	346.8	30.3	458	13 BX668737	BX668737 BX668737
21	323.6	28.2	552	9 AV599392	AV599392 AV599392
22	316.8	27.6	670	10 BB649319	BB649319 BB649319
23	284.6	24.8	463	10 AW744257	AW744257 ur30a06.Y
24	277.2	24.2	654	13 BU310812	BU310812 603541586
25	277.2	24.2	654	13 BU312228	BU312228 603544226
26	267.2	23.3	448	12 BG793104	BG793104 UTSW H8E1
27	248.4	21.7	404	9 AI414584	AI414584 ma51d03.X
28	243	21.2	410	9 AV599391	AV599391 AV599391
29	236	20.6	344	14 W09785	W09785 ma51d03.Y
30	229.8	20.1	624	13 BX740493	BX740493 BX740493
31	219.8	19.2	432	9 AA039033	AA039033 ml97a10.Y
32	218.6	19.1	401	13 BQ211290	BQ211290 UI-R-DY1-
33	211	18.4	974	29 CNS03MEA	AL250651 Tetraodon
34	198.2	17.3	368	9 AI172339	AI172339 BST218341
35	198	17.3	779	14 CA969408	CA969408 CCLX06822
36	193.4	16.9	552	13 BX869477	BX869477 BX869477
37	193.4	16.9	695	13 BX869001	BX869001 BX869001
38	193.4	16.9	728	13 BX869706	BX869706 BX869706
39	193.4	16.9	744	13 BX871943	BX871943 BX871943
40	193.4	16.9	756	13 BX868214	BX868214 BX868214
41	184	16.1	401	14 W36976	W36976 mb82f10.Y
42	183	16.0	275	14 W18106	W18106 mb84f07.Y
43	176	15.4	367	13 BY340299	BY340299 BY340299
44	175	15.3	329	9 AI716006	AI716006 UI-R-Y0-a
45	171.2	14.9	359	13 BY341219	BY341219 BY341219

ALIGNMENTS

RESULT 1

AK081696

LOCUS

DEFINITION

Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone: C130080I2 product: weakly similar to F36F2.1 PROTEIN [Caenorhabditis elegans], full insert sequence.

1763 bp mRNA linear HTC 20-SEP-2003

ACCESSION

AK081696

VERSION

AK081696.1 GI:26349314

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus

ORGANISM

Mus musculus

REFERENCE

1

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

JOURNAL

92727253

MEDLINE

10349636

PUBMED

11042159

2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

2049374

MEDLINE

PUBMED

QY 901 COTGCTGAGGACACATCTACAGGAATGATGGACATGTGCTTCAATTCATGCAATG 960
 Db 936 CGAGCGGAGAGACACATCTATCGGAAATATGGAACATGTGCTTGTATCGGACAAATG 995
 QY 961 GTCGCCACAGACGAGATGGCAAGATCCAGGTTACTTTTGAGATCTCTTTGACAGATAC 1020
 Db 996 GTCGCCACAGACGAGATGGCAAGATCCAGGTTACTTTTGAGATCTCTTTGATCGGTAT 1055
 QY 1021 GTCTGATTTTCAGATAAGTAGTGGCAATTCATCGGTGCCAGGAACATGGACTGTA 1080
 Db 1056 GTTCGAATTTCTGATAAAGTCGTGGGCATCTCTCATCGTCCGAGAAACACGGACTGGT 1115
 QY 1081 GACTTTGAAGGAGAGATGCTATGGCAAGCCGAGATGACCATGTTGTGATTACGCTACTC 1140
 Db 1116 CACTTTGAAGGAGAGATGCTATGGCAAGCCGAGACCATGTTGTGATTACTCTCTT 1175
 QY 1141 AAGTGA 1146
 Db 1176 GAGTAA 1181

AK045186 2149 bp mRNA linear HTC 20-SEP-2003
 Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
 enriched library, clone.B130044G13 product:weakly similar to
 F36P2.1 PROTEIN [Caenorhabditis elegans], full insert sequence.

ACCESSION AK045186
 VERSION AK045186.1 GI:26337134
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

REFERENCE 2
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20493374
 PUBMED 11042159

REFERENCE 3
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kutsunai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanishi, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861

REFERENCE 4
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 REFERENCE 5
 AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 REFERENCE 6
 (bases 1 to 2149)

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.

TITLE

JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by Dr. Tomohiro Kono (Department of Animal
 Science, Tokyo University of Agriculture, 1737 Hunko Atsugi City,
 Kanagawa Prefecture, Japan) whose assistance we gratefully
 acknowledge.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/.

FEATURES

source

1. 2149
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM.DB:B130044G13"
 /db_xref="MGI:2410542"
 /db_xref="taxon:10090"
 /clone="B130044G13"
 /tissue_type="parthenogenote"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="9.5 days embryo"
 <1..1009

CDS

/note="unnamed protein product; putative
 weakly similar to F36P2.1 PROTEIN [Caenorhabditis elegans]
 (SPTT|062234, evidence: FASTA, 42.3%ID, 84.5%length,
 match=423)"
 /codon_start=2
 /protein_id="BAC32252.1"
 /db_xref="GI:26337135"

ORIGIN

Query Match 60.8%; Score 636.8; DB 11; Length 2149;
 Best Local Similarity 81.0%; Pred. No. 1.2e-166;
 Matches 831; Conservative 0; Mismatches 177; Indels 18; Gaps 1;
 QY 121 AGCATCAGCGAGCCAGGAGCTACAGCTGGCTGGCGGGAGGAGCCAGGACTCCT 180
 Db 2 AGTACCAACAGCCAGGAGCTCAGCTGGCTGGCGGGAGGAGCACTCATGACCTACT 61
 QY 181 CAAGCTCTTAAACCAATCACACCCCTACTTCCACACCAGAAAGCTCAGAGTCCCAAG 240
 Db 62 AAGCTCTTAAAGAGAGCGGCTCTTAC-----CAGCATGCCCCCAAA 103


```

Db      541 CATGACTGTAGACTTTTGAAGGAGAGATGCTATGCGAAGCCGAGATGACCATGTTGTG 600
QY      1129 ATTACGCTACTCAAGTGA 1146
Db      601 ATTACGCTACTCAAGTGA 618

RESULT 4
LOCUS   AY399109
DEFINITION Pan troglodytes HCM0114 gene, DNA linear GSS 12-DEC-2003
genomic survey sequence.
ACCESSION AY399109
VERSION   1
KEYWORDS  GSS.
SOURCE    Pan troglodytes (chimpanzee)
ORGANISM  Pan troglodytes
REFERENCE 1 (bases 1 to 618)
AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
          Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,
          Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
          Adams,M.D. and Cargill,M.
TITLE      Inferring nonneutral evolution from human-chimp-mouse orthologous
          gene trios
JOURNAL    Science 302 (5652), 1960-1963 (2003)
PUBMED    14671302
REFERENCE 2 (bases 1 to 618)
AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
          Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,
          Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
          Adams,M.D. and Cargill,M.
TITLE      Direct Submission
JOURNAL    Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
          Rockville, MD 20850, USA
COMMENT    This sequence was made by sequencing genomic exons and ordering
          them based on alignment
FEATURES   Location/Qualifiers
            source          1..618
                        /organism="Pan troglodytes"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9598"
            gene            <1..>618
                        /locus_tag="HCM0114"
ORIGIN
Query Match 53.8%; Score 616.4; DB 29; Length 618;
Best Local Similarity 99.8%; Pred. No. 2e-146;
Matches 617; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      529 ATGAGCAGGAGGAGCCCATGAGGAGTGACAGCGGTAGACACAGAGGACAGCGGCTAT 588
Db      1 ATGAGCAGGAGGAGGAGCCCATGAGGAGTGACAGCGGTAGACACAGAGGAGCAGCGGCTAT 60
QY      589 GGAGGAGAGGCTGAGGAGAGCCCGCAGCAGGATGGAGTGCGAGTGCGTGTGGTCAGGATC 648
Db      61 GGAGGAGAGGCTGAGGAGAGCCCGCAGCAGGATGGAGTGCGAGTGCGTGTGGTCAGGATC 120
QY      649 AAGGCCCTTCCCTCCAGGTAAACAGATTTACAGAACTCACTCAAGCCCA 708
Db      121 AAGGCCCTTCCCTCCAGGTAAACAGATTTACAGAACTCACTCAAGCCCA 180
QY      709 CAGAAATATAGCCAGTGGGCAACTTTAAAGGAGATGGCAGCAGTGCGGTGTGATGAACAC 768
Db      181 CAGAAATATAGCCAGTGGGCAACTTTAAAGGAGATGGCAGCAGTGCGGTGTGATGAACAC 240
QY      769 ATACAATCCCAAGAGCTCAATCTTTTCAGTGAAGATTTGATTACGAGCTGCCATGCC 828
Db      241 ATACAATCCCAAGAGCTCAATCTTTTCAGTGAAGATTTGATTACGAGCTGCCATGCC 300
QY      829 ACCGCCCTACAAAGGAGATGAGGGCTATGCGCGCCCCCAAGAGGAACCAAACTGCT 888

```

```

Db      301 ACCCGCTACAAAGGAGATGAGGGCTATGCGCGCCCCCAAGAGGAACCAAACTGCT 360
QY      889 GAAAGGCCCAAGCGTCTCGAGGAGCACATCTACAGGAAATGATGACATGCTTCATT 948
Db      361 GAAAGGCCCAAGCGTCTCGAGGAGCACATCTACAGGAAATGATGACATGCTTCATT 420
QY      949 ATCTGCACATGCTCGCCACAGACAGATGGCAAGATCCAGGTTACTTTTGGAGATCTC 1008
Db      421 ATCCGACAAATGCTCGCCACAGACAGATGGCAAGATCCAGGTTACTTTTGGAGATCTC 480
QY      1009 TTTGACAGATACGTTGCTATTTTTCAGATAAAGTAGTGGGCAATTCATGCGTCCAGGAAA 1068
Db      481 TTTGACAGATACGTTGCTATTTTTCAGATAAAGTAGTGGGCAATTCATGCGTCCAGGAAA 540
QY      1069 CATGGAAGTGTAGACTTTTGAAGGAGATGCTATGCGAGCGGAGATGACCATGTTGTG 1128
Db      541 CATGGAAGTGTAGACTTTTGAAGGAGATGCTATGCGAGCGGAGATGACCATGTTGTG 600
QY      1129 ATTACGCTACTCAAGTGA 1146
Db      601 ATTACGCTACTCAAGTGA 618

RESULT 5
LOCUS   BF670302
DEFINITION 602149566F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4290760 5',
mRNA sequence.
ACCESSION BF670302
VERSION   1
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 896)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
          National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-remail.nih.gov
          Tissue Procurement: CLONETECH Laboratories, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCM134 row: C column: 17
          High quality sequence stop: 525.
          Location/Qualifiers
            source          1..896
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:4290760"
                        /lab_host="NIH_MGC_81"
                        /note="Organ: muscle (skeletal); Vector: pDNR-LIB
          (Clontech); Site 1: SfiI (ggccgcttcggcc); Site 2: SfiI
          (ggccatcatggcc); 5' and 3' adaptors were used in cloning
          as follows: 5' adaptor sequence: 5'-CACGCCCATTTATGGCC-3'
          and 3' adaptor sequence:
          5'-ATTCTAGAGCGCGCGCCGACATG-dT(30)BN-3' (where B = A,
          C, or G and N = A, C, G, or T). Average insert size
          1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
          inserts by PCR. This library was enriched for full-length
          clones and was constructed by Clontech Laboratories (Palo
          Alto, CA)."
ORIGIN
Query Match 43.7%; Score 501.2; DB 10; Length 896;

```

```

Best Local Similarity 98.9%; Pred. No. 7.2e-117;
Matches 526; Conservative 0; Mismatches 3; Indels 3; Gaps 2;

QY 618 GGATGGAGTGCAGGTGGCTGGTTCAGATCAAGCGCCCTTGCCTCCAGGTAAACAG 677
Db 1 CGATGGAGTGCAGGTGGCTGGTTCAGATCAAGCGCCCTTGCCTCCAGGTAAACAG 60
QY 678 ATTTACAGAGAACTCAACTGCAAGGCGCCCAAGAAATATAGCCAGTGGGCACTTGAA 737
Db 61 ATTTACAGAGAACTCAACTGCAAGGCGCCCAAGAAATATATAGCCAGTGGGCACTTGAA 120
QY 738 AGGAGATGGGAGAGTGGCTGTATGACACATCAATCCAGAGAGTCAATCCCTTTTCA 797
Db 121 AGGAGATGGGAGAGTGGCTGTATGACACATCAATCCAGAGAGTCAATCCCTTTTCA 180
QY 798 TGAAGAGTTTGATTCAGAGTGGCCATCTCCACCGCTTACAAAGAGAGATGAGGCTA 857
Db 181 TGAAGAGTTTGATTCAGAGTGGCCATCTCCACCGCTTACAAAGAGAGATGAGGCTA 240
QY 858 TGGCGCGCCCAAGAGAGNACCAAACTGCTGMAAGGCGCAAGCTGCTGAGGAGCACAT 917
Db 241 TGGCGCGCCCAAGAGAGNACCAAACTGCTGMAAGGCGCAAGCTGCTGAGGAGCACAT 300
QY 918 CTACAGGGAATATGATGGACATGCTTCAATATCTGCAATATGCTGCTGCCACAGAGA 977
Db 301 CTACAGGGAATATGATGGACATGCTTCAATATCTGCAATATGCTGCTGCCACAGAGA 360
QY 978 TGGAGATCCAGGTACTTTTGGAGATCTTTTGCAGATACGTTCTGTTTTCAGAT-A 1036
Db 361 TGGAGATCCAGGTACTTTTGGAGATCTTTTGCAGATACGTTCTGTTTTCAGAT-A 420
QY 1037 AAGTAGTGGCCATCTTCATCGCTGCCAGGAAACATGGAGTGGTAGACTTTGAAGAGAGA 1096
Db 421 AAGTAGTGGCCATCTTCATCGCTGCCAGGAAACATGGAGTGGTAGACTTTGAAGAGAGA 480
QY 1097 TGTATGGCAAGG--CCGAGATGACCATGTTGATTCAGTACTCAAGTGA 1146
Db 481 TGTATGGCAAGGCGCGAGGATGACCATGTTGATTCAGTACTCAAGTGA 532

RESULT 6
BX671002
LOCUS
DEFINITION
BX671002 Sus Scrofa library (scac) Sus scrofa cDNA clone
scac00351.h.22 3prim, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
EST.
GI:37982096
Sus scrofa (pig)
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 717)
Bonnat, A., Tosser-Klopp, G., Benne, F., Cabau, C., Villegier, S.,
Soares, M., Bonaldo, F. and Haeve, F.
A Pig Normalised Multi-Tissue cDNA Library
Unpublished (2003)
Contact: Tosser-Klopp G
Genetique Animale
Institut National de la Recherche Agronomique
Chemin de Borde-Rouge - Auzeville BP27, 31326 Castanet-Tolosan
cedex, FRANCE
Tel: 33 (0) 5.61.28.51.14
Fax: 33 (0) 5.61.28.53.08
Email: tosse@toulouse.inra.fr
Clone distribution: AGENAB Resource centre, Francois PIUMI,
Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et Etude du
Genome (LURE), Domaine de Vilvert, 78352, Jouy-en-Josas cedex,
FRANCE, +33 (0) 1.34.65.28.02, +33 (0) 1.34.65.22.73
Sequence cleaned of vector, adaptor and repetitions. Contact us
at sigenasupport@jouy.inra.fr to obtain the chromatogram of this
sequence.
Plate: 0035 row: h column: 22.

```

FEATURES

source

Location/Qualifiers

1..717
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/clone="scac00351.h.22"
/tissue_type="mixed"
/clone_lib="Sus Scrofa library (scac)"
/note="Vector: pT73D-pac vector; tissues: adipose tissue, brain, kidney, liver, muscle, ovary, testis, heart, hypothalamus, pancreas, skin, spleen, thymus, placenta, pituitary gland, seminal vesicle, small intestine, uterus, adrenals, bulbo urethral gland, cerebral trunk, epididymis, female gonad, gall-bladder, hippocampus, large intestine, male gonad, melanocytes, stomach, udder"

ORIGIN

Query Match 43.2%; Score 494.8; DB 13; Length 717;
Best Local Similarity 89.6%; Pred. No. 2.7e-115;
Matches 532; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 553 AGGAGTGACAGCGTAGACACAGAGGACAGCGCTATGGAGGAGAGGCTCAGAGAGGCCC 612
Db 1 AGGAGTGATAGCTATAGACACAGAGGACAGCGCTACGGAGGGGAGACAGAGAGGCGC 60
QY 613 GAGCAGGATGGAGTGCAGGTGGCTGPGTCAAGATCAAGCGCCCTTGCCTCCAGGTA 672
Db 61 GAGCAGGATGGAGGAGGAGGCGCATCGCCAGAAATCAAAACGCGCCCTTGCCTCCAGGCA 120
QY 673 AACAGATTACAGAGAACTCACTGCAAGGCGCCACAGAAATATAGCCAGTGGGCAAC 732
Db 121 AATAGATTATACCGAGAACTCAATTCGAAAGCCCGGGAATATACGCAAGTGGGCGCAC 180
QY 733 TTGAAAGGGAGATGGCAGCAGTGGGCTGATGAACACATCAATCCAGAGGCTCAATCCT 792
Db 181 CTGAAAGGGAGGTGGCAGCAGTGGGCTGATGAACACATCAATCCAGAGGCTCAACCT 240
QY 793 TTGAGTGAAGATTGATTTAGCAGCTGGCCATGTCACCGCTACACAAAGGAGATGAG 852
Db 241 TTCAGTGACGAGTTGATTATGAAATTTGGCCATGTCACCGCTTGCACAAAGGAGATGA 300
QY 853 GGCTATGGCGCCGCCAAAGAGAAACCAAACTGCTGAAAGGCGCCAGCGTGTGAGGAG 912
Db 301 GGCTACGCGCGTCCCAAGGAGAAACCAAGAACTGCGGAAAGGCTAAGAGAGTGAAG 360
QY 913 CACATCTACAGGGAATGATGGACATGCTTCAATATCTGCACATGCTCCGCCACAGA 972
Db 361 CACATCTACAGAGAAATCATGGACATGCTTCAATCATCCGACGATGGCGCACCCAGG 420
QY 973 CGAGATGGCAAGATCCAGGTTACTTTTGGAGATCTCTTTGACAGATACGTTCTGATTTC 1032
Db 421 CGGACCGCAAGATCCAGGTTACTTTGAGAGATCTCTTTGACAGATACGTTCTGATTTC 480
QY 1033 GATAAAGTAGTGGGCACTTCTCATCGTGGCCAGGAAACATGGACTGGTAGACTTTGAGGA 1092
Db 481 GATAAAGTAGTGGGCACTTCTCATCGCGCCAGGAAACACGAGTGGTGTGACTTTGAGGA 540
QY 1093 GAGATGCTATGGCAAGCGCCAGATGACCATGTTGATTTACGCTACTCTCAAGTGA 1146
Db 541 GAGATGCTATGGCAAGCGCCAGATGACCATGTTGATTTACGCTACTCTGCTAAAGTGA 594

RESULT 7

BX500732

LOCUS

DEFINITION

DKZP779K1758.1 779 (synonym: hnccl) Homo sapiens cDNA clone

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 537)
 Koehrer, K., Beyer, A., Mewes, H.W., Weil, B., Amid, C., Osanger, A.,
 Fobo, G., Han, M. and Wienann, S.
 EST (Koehrer, K., Beyer, A., Mewes, H.W., Weil, B., Amid, C., et al.)
 Unpublished (2003)
 Contact: MIPS
 MIPS
 Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by BWFZ (Biomedical Research Center at the Heinrich-
 Heine-University, Duesseldorf/Germany) within the cDNA sequencing
 consortium of the German Genome Project. No sl sequence available.
 This clone (DKFZp779K1758) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

source
 1..537
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFZp779K1758"
 /tissue_type="liver"
 /dev_stage="fetal"
 /lab_host="DH10B"
 /clone_lib="779 (synonym: hmccl)"
 /notes="Vector: pSport1_Sfi; Site 1: SfiIA; Site 2: SfiIB"

ORIGIN

Query Match 42.0%; Score 481.4; DB 13; Length 537;
 Best Local Similarity 99.8%; Pred. No. 6.1e-112;
 Matches 482; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ATGGCTCCGGGCGAAAGAAAGCGGGAGGGCCAGCAAGAGCGCCCTCCGGAGATA 60
 Db 55 ATGGCTCCGGGCGAAAGAAAGCGGGAGGGCCAGCAAGAGCGCCCTCCGGAGATA 114
 QY 61 CGCACAGCCACCTGTCTATCAGCTTGGCCGAGGTTGGCAGAGTGGCGGATGAGAC 120
 Db 115 CGCACAGCCACCTGTCTATCAGCTTGGCCGAGGTTGGCAGAGTGGCGGATGAGAC 174
 QY 121 AGCATCAGGCGAGGCCAGGAGCCTACAGGCTGGCTGCCGGAGGGACCCAGGACTCACT 180
 Db 175 AGCATCAGGCGAGGCCAGGAGCCTACAGGCTGGCTGCCGGAGGGACCCAGGACTCACT 234
 QY 181 CAAGTCTTAACCAATCAACCCCTCTATTCTACACAGAAAGCTCAGAGTGGCCCAAG 240
 Db 235 CAAGTCTTAACCAATCAACCCCTCTATTCTACACAGAAAGCTCAGAGTGGCCCAAG 294
 QY 241 TCGCCACCCCGCTGCCAGAGGACATGGAGATGGACAAAGCTTCAGAGAAAGCCCTGAG 300
 Db 295 TCGCCACCCCGCTGCCAGAGGACATGGAGATGGACAAAGCTTCAGAGAAAGCCCTGAG 354
 QY 301 GTTTCTCAATCAATAAAGAGAGGTTCATAACGGTGTCTAGCAAGACTTACGAGAGA 360
 Db 355 GTTTCTCAATCAATAAAGAGAGGTTCATAACGGTGTCTAGCAAGACTTATGAGAGA 414
 QY 361 GGAGGGAGCTGAGCCACCTCAGCCAGGTTACGAGAGGATGCTGTGTGCTTGAACCT 420
 Db 415 GGAGGGAGCTGAGCCACCTCAGCCAGGTTACGAGAGGATGCTGTGTGCTTGAACCT 474
 QY 421 GGGGAGCCAGAGATGACATTGACAGAAATCTTCCAGAGCCACCGCTCCCAACCGGAGG 480
 Db 475 GGGGAGCCAGAGATGACATTGACAGAAATCTTCCAGAGCCACCGCTCCCAACCGGAGG 534
 QY 481 AGA 483
 Db 535 AGA 537

RESULT 8

AY399110
LOCUS

618 bp DNA linear GSS 12-DEC-2003
 Mus musculus HCM0114 gene, VIRTUAL TRANSCRIPT, partial sequence,
 genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AY399110.1 GI:39755099
 GSS.
 Mus musculus (house mouse)
 Mus musculus
 Mus musculus

REFERENCE
AUTHORS

1 (bases 1 to 618)
 Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
 Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
 Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
 Adams, M.D. and Cargill, M.

TITLE

Inferred nonneutral evolution from human-chimp-mouse orthologous
 gene trios

JOURNAL
PUBMED

Science 302 (5652), 1960-1963 (2003)
 14671302

REFERENCE
AUTHORS

2 (bases 1 to 618)
 Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
 Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
 Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
 Adams, M.D. and Cargill, M.

TITLE
JOURNAL

Direct Submission
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA

COMMENT

This sequence was made by sequencing genomic exons and ordering
 them based on alignment.

FEATURES
source

1..618
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /locus_tag="HCM0114"

ORIGIN

Query Match 41.4%; Score 474; DB 29; Length 618;
 Best Local Similarity 85.4%; Pred. No. 5.2e-110;
 Matches 528; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
 QY 529 ATGAGCAGGAGGAGGCCACATGGAGGATGACAGCTAGACAGAGGAGCGGCTAT 588
 Db 1 ATGGAACAGGAAGAGCCACGCTGGAAGAGTACAGCGGTAGACAGAGGAGCGGCTAT 60
 QY 589 GGAGGAGAGGCTGAGGAGAGGCCGAGCAGGATGGAGTGCAGGTGGCTGGTCAGGATC 648
 Db 61 GGAGGAGATGAGGAGAGGCCGAGCAGGATGGAGTGCAGGTGGCTGGTCAGGATC 120
 QY 649 AAGGCCCTTGGCCCTCCAGGTAAAGAGATTACAGAGAACTCACTGCAAGCCCAA 708
 Db 121 AAACGCCCTTGGCTCTCCAGGCAACAGGTACTCTGAGCCACTCAACTGTAAGCCCAT 180
 QY 709 CAGAAATATAGCCAGTGGCAACTTGAAGAGGAGATGGCAGCAGTGGGCTGATGAACAC 768
 Db 181 CGAAATACGCCAAGTGGCAACTTGAAGAGGAGTGGCAGCAGTGGGCTGATGAACAC 240
 QY 769 ATACAATCCCAAGAGCTCAATCCTTTAGTGAAGAGTTTGTATTAGAGCTGCCATGTCC 828
 Db 241 GTCCAGTCCCAAGAGCTCAATCCTTTAGTGAAGAGTTTGTATTAGAGCTGCCATGTCC 300
 QY 829 ACCCGCTTACCAAGAGGATGGGCTATGGCGCCCAAGAGGAGCAACAAACTGCT 888
 Db 301 ACTCGGCTCCCAAGGAGAGCGAGGCTATGGCGCCCAAGAGGAGCAACAGACGCT 360
 QY 889 GAAAGGGCCAGCGTGTCTGAGAGCACATCTACAGGGAATGATGACATGTGCTTACT 948
 Db 361 GAAAGGGCCAGCGGAGAGAGAGCACATCTATCGGAAATATGGAATGTGCTTGTGT 420
 QY 949 ATCTGCACATGCTCGCCACAGACAGATGGCAGATCCAGGTTACTTTTGAGATCTC 1008

Db 421 ATCCGACATGGCTCGCCACAGACGAGATGGCAAGATCCAGTTACTTTTCGGAGAACTC 480
 QY 1009 TTTGACAGATCGTTCGATTTCAGATAAAGTACTGGCATTCTCATGGCGGCCAGAAA 1068
 Db 481 TTTGATCCTATGTTCCGATTCTTGATAAAGTCGTGGCATCTTCATCGGTGCCAGAAA 540
 QY 1069 CATGACTGGTAGACTTTTGAAGGAGAGATGCTATGGCAAGCCGAGATGACCAATGTTGTG 1128
 Db 541 CACGACTGGTGCATCTTTGAAGGAGAGATGCTATGGCAAGCCGAGACGACCATGTTGTG 600
 QY 1129 ATTACGCTACTCAAGTGA 1146
 Db 601 ATTACTCTCGTTGAGTAA 618

RESULT 9
 LOCUS AI605651 576 bp mRNA linear EST 15-MAR-2000
 DEFINITION mas1d03.y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
 IMAGE:314213 5' similar to WP:F36F2.1 CB15975 ;, mRNA sequence.

ACCESSION AI605651
 VERSION AI605651.1 GI:4614818
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; 1 (bases 1 to 576)

REFERENCE 1 (bases 1 to 576)
 AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterson,R. and Wilson,R.
 TITLE The WashU-NCI Mouse EST Project 1999
 JOURNAL Unpublished (1999)

COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
 This read is a RESEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the correct orientation)
 Possible reversed clone: similarity on wrong strand

NGI:204829
 Seq primer: -40RP from Gibco
 High quality sequence stop: 466
 POLYA=No.

FEATURES
 source Location/Qualifiers
 1..576
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:314213"
 /dev_stage="19.5 dpc total fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares mouse p3NMF19.5"
 /note="Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer 15' TGTACCAATCGAATGGAGCGCGCATTTTTTTTTTTTTTTT 3', double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

ORIGIN

Query Match 37.4%; Score 429; DB 9; Length 576;
 Best Local Similarity 83.9%; Pred. No. 1.6e-98;
 Matches 483; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
 QY 406 GGTGTCTTGAACCTGGGCGAGGACGATGACATTCAGATGATCCACAGCCACGCG 465
 Db 1 GCGGTGTCTGGAAGCTATTTCAGCCAGAGATGACATTCAGATGATCCCTTTAGTCACGAC 60
 QY 466 TCCCAACCGCGGAGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGCTGGAGA 525
 Db 61 TCGCCAAACGCGAGAGAAAATGCACCAACCTGGTGTCTGAGCTGCCAAAGGCTGGAAA 120
 QY 526 GTGATGACGAGGAGGCCCATGCTGAGGAGTACAGCTAGACAGCTAGACACAGAGACGCGC 585
 Db 121 GTGATGGAACAGGAAGAGCCCGTGGAAAGATGACAGCTAGACAGAGACAGTGGC 180
 QY 586 TATGGAGGAGAGGCTTGGAGAGGCGCCGAGCAGGATGGAGTGCAGTGGCTGTGTGTCAGG 645
 Db 181 TACGGAGGGGATATGGAGGAGAGGCTGTGAGCAAGATGACAGCGCTGTGGCTCTCTGCCAGG 240
 QY 646 ATCAAGCGCCCTTGGCCCTCCAGAGTAAACAGATTTACAGAGAAATCAACTGCAAGGCG 705
 Db 241 ATCAAAACGCGCCCTTGTCTCTCCAGGCAACAGGTACTCTGAGACACTCAACTGTGAAGGCC 300
 QY 706 CACACAGAAATATAGCCAGTGGGCAACTTTGAAAGGAGATGGCAGCAGTGGGCTGATGAA 765
 Db 301 CATCGGAATACAGCCAACTGGACAACTTTGAAAGGAGGCTGGCAGCAGTGGGCGCATGAA 360
 QY 766 CACATCAATTCACAGAAAGTCAATCTTTTCAGTGAAGATTTGATTACAGCTGGCCATG 825
 Db 361 CAGCTCAGTCCCAAGAGCTCAATCCCTTTTCAGTGAAGATTTGATTACAGCTGGCCATG 420
 QY 826 TCCACCGGCTACACAAAGGAGATGAGGCTTATGGCGCCCAAGAGAGAAACCAAACT 885
 Db 421 TCCACTCGGCTCCAAAGGAGACAGAGGCTATGGCGCCCAAGAGAGAAACCAAGAC 480
 QY 886 GTGAAAGGCGCCAAAGCGTGTGAGGAGACATCTACAGGAAATGATGACATGCTCTTC 945
 Db 481 GCNTANAGGGCCAAAGCGAGCGGAAGAGCACATCTATCGGAAATTTATGGAACCTGCTTNT 540
 QY 946 ATTATCTGCACATGGCTGCCACACAGCAGATGGC 981
 Db 541 GTTATCGGCAATGGCTGCCACACAGCAGATGGC 576

RESULT 10

LOCUS BX667447 909 bp mRNA linear EST 22-OCT-2003
 DEFINITION BX667447 Sus Scrofa library (scag) Sus scrofa cDNA clone
 scag0001c.b.12 3prim, mRNA sequence.

ACCESSION BX667447
 VERSION BX667447.1 GI:37857165
 KEYWORDS EST.
 SOURCE Sus scrofa (pig)

ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus; 1 (bases 1 to 909)

REFERENCE 1 (bases 1 to 909)
 AUTHORS Bonnet,A., Tosser-Klopp,G., Benne,F., Cabau,C., Villegier,S., Soares,M., Bonaldo,F. and Hatay,F.
 TITLE A pig Normalised Multi-Tissue cDNA Library

COMMENT Unpublished (2003)
 Contact: Tosser-Klopp G
 Genetique Animale
 Institut National de la Recherche Agronomique
 Chemin de Borde-Rouge - Auzeville BP27, 31326 Castanet-Tolosan
 cedex, FRANCE

Tel: 33 (0) 5.61.28.51.14
 Fax: 33 (0) 5.61.28.53.08
 Email: tosser@coulose.inra.fr
 Clone distribution: AGENAE Resource centre. Francois PIUMI,
 Francois.PIUMI@jouy.inra.fr, INRA, CEA Radiobiologie et Etude du
 genome (LREG), Domaine de Vilvert, 78352, Jouy-en-Josas cedex,

genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.

FEATURES

Location/Qualifiers
 1. .684
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="D330042E04"
 /tissue_type="heart"
 /dev_stages="13 days embryo"
 /lab_host="DH10B"
 /clone_lib="RIKEN full-length enriched, 13 days embryo
 heart"

/note="Site 1: SalI; Site 2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGCGCGCAACTCGAGTGTGTTTTTTTTTTTNN 3'], cDNA was
 prepared by using trihalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. Second strand cDNA was prepared with the
 primer adapter of sequence [5'
 GAGAGAGAGTCTCGAGTGAATTAATTAATTCCTCCCCCCCC 3']. cDNA
 was cleaved with BamHI and XhoI. Vector: a modified
 pBluescript KS(+) after bulk excision from Lambda FLC 1."

ORIGIN

Query Match 37.0%; Score 423.8; DB 13; Length 684;
 Best Local Similarity 77.6%; Pred. No. 3.8e-97;
 Matches 545; Conservative 0; Mismatches 138; Indels 19; Gaps 2;
 QY 121 AGCATCAGGAGCCAGGAGCTACAGCTGGCTGCCGGAGGAGCCAGGACTCAGCT 180
 Db 2 AGTACCAACAGCCAGGAGCTGCAGCTGGCTGCCGGAGCAACTCATGACCTACT 61
 QY 181 CAAGCTCTTAACCAATCACACCCCTACTTCACACCAGAAAGCTCAGAGTCCCCCAAG 240
 Db 62 AACGCTCCTAAAGAGCCGGTCTCTTAC-----CAGCATCCCCCAGAA 103
 QY 241 TGCCACCCCGCTGCCAGAGGACATGGAGTGGACAAAGCTCAGAGAAAGCCCTGAG 300
 Db 104 ACTCTGTCTCAAGCCAGATCGAGAGGAGGACACACTCAGAAAGCCACCGAG 163
 QY 301 GTTCTTCACATCAAAAGAAAGAGGTGTCAAAACCGTGTGTCAGGAAAGCTTACGAGAGA 360
 Db 164 GTCTCCACATTAAGAAAGAGAGGTGACAGAAACGGTTGTTCAGCAAGGCTTATGAGAGG 223
 QY 361 GGAGGGAGCTGAGCCACCTCAGCCAGATCAGAGGAGTCTCGTGTGTTGAACCT 420
 Db 224 GGAGGAGATGTGAATCTCTGAGCCACAGGTATGAAATGATGGCGGCTGTCTGAGCT 283
 QY 421 GGGCAGCCAGAGATGACATTGACAGATCTCCACAGCCACCGCTCCCAACCGGGAGG 480
 Db 284 ATTCAAGCCAGAGATGACATTGACAGATCTCTTAGTCACGACTCGCAACACCGAGA 343
 QY 481 AGAAATGTCACCACTGGTCTCTGAGCTAACCAAGGCTGGAGAGTGTGAGCAGGAG 540
 Db 344 AGAAATGTCACCACTGGTCTCTGAGCTGACCAAGGCTGGAAAGTGTGTAACAGGAA 403

QY 541 GAGCCCATGAGGAGGTGACAGCTAGACACAGAGGACAGCGGTATGAGGAGAGGCT 600
 Db 404 GAGCCCATGAGGAGGTGACAGCTAGACACAGAGGACAGCGGTATGAGGAGAGGCT 463
 QY 601 GAGGAGAGCCCGAGCAGGATGAGGAGTGGAGTGGCTGTGCTCAGGATCAAGCGCCCTTG 660
 Db 464 GAGGAGAGCCCTGAGCAGATGAGGAGTGGCTGTGCTCAGGATCAAGCGCCCTTG 523
 QY 661 CCTCCAGGTAACAGATTTACAGAGAACTCACTGCAAGCCCAACAGAAATATAGC 720
 Db 524 CTCCTCCAGGCAACAGAGTACTCTGAGACACTCAACTGTAAAGGCCCATC-GGAATACAGC 582
 QY 721 CCAGTGGGCAACTTGAAGGGAGATGGCAGCAGTGGCTGTGATGAACACATCAATCCCAAG 780
 Db 583 CAACTGGACAACTTGAAGGGAGTGGCAGCAGTGGCCGATGAACACGTCAGTCCCAAG 642
 QY 781 AAGCTCAATCCTTTTCACTGTAAGAGTTTGAATACGAGCTGGCC 822
 Db 643 AAGCTCAATCCTTTTCACTGTAAGAGTTTGAATACGAGCTGGCC 684

RESULT 12

EX675669 624 bp mRNA linear EST 28-OCT-2003
 LOCUS BX675669 Sus Scrofa library (scac) Sus scrofa cDNA clone
 DEFINITION scac00301.n.16 3prim, mRNA sequence.

ACCESSION BX675669
 VERSION BX675669.1 GI:38009804

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Sus scrofa (pig)
 Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 624)
 Bonnet A., Tosser-Klopp G., Benne F., Cabau C., Villegier S.,
 Soares M., Bonaldo F. and Haeve F.,
 A Pig Normalised Multi-Tissue cDNA Library
 Unpublished (2003)
 Contact: Tosser-Klopp G
 Genetique Animale
 Institut National de la Recherche Agronomique
 Chemin de Borde-Rouge - Auzeville BP27, 31326 Castanet-Tolosan
 cedex, FRANCE
 Tel: 33 (0) 5.61.28.51.14
 Fax: 33 (0) 5.61.28.53.08
 Email: tosser@clouse.inra.fr

Clone distribution: AGENAS Resource centre, Francois PIUMI,
 Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et Etude du
 Genome (LREG), Domaine de Vilvert, 78352, Jouy-en-Josas cedex,
 FRANCE, +33 (0) 1.34.65.28.02, +33 (0) 1.34.65.22.73
 Sequence cleaned of vector, adaptor and repetitions. Contact us
 at sigenasupport@jouy.inra.fr to obtain the chromatogram of this
 sequence.

Plate: 0030, row: n column: 16.

FEATURES

source

Location/Qualifiers
 1. .624
 /organism="Sus scrofa"
 /mol_type="mRNA"
 /db_xref="taxon:9823"
 /clone="scac00301.n.16"
 /tissue_type="mixed"
 /clone_lib="Sus Scrofa library (scac)"
 /note="Vector: p7T3D-pac vector; tissues: adipose tissue,
 brain, kidney, liver, muscle, ovary, testis, heart,
 hypothalamus, pancreas, skin, spleen, thymus, placenta,
 pituitary gland, seminal vesicle, small intestine,
 uterus, adrenals, bulbo urethral gland, cerebral trunk,
 epididymis, female gonad, gall-bladder, hippocampus,
 large intestine, male gonad, melanocytes, stomach, udder"

ORIGIN

Query Match 36.5%; Score 418.2; DB 13; Length 624;

```

Best Local Similarity 90.3%; Pred. No. 9.7e-96;
Matches 447; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 652 CGCCCTTCCCTCCAGTAAACAGATTTCAGAGAACTCACTGCAAGCCCAACAG 711
Db 1 CGCCCTTCCCTCCAGCAATAAGATTTCAGAGAACTCAATTCGAAGCCCAAGCGG 60
QY 712 AAATATAGCCCTAGTGGCAACTTGAAGGGAGATGGCAGCAGTGGGCTGATGAACACATA 771
Db 61 AAATACAGCAAGTGGCCACCTGAAGGGAGGTGGCAGCAGTGGGCTGATGAACACATA 120
QY 772 CAATCCAGAGCTCAATCTTTCACTGAAGAGTTTGATTAGACGTGGCCATGTCACC 831
Db 121 CAATCCAGAGCTCAACCTTTTCACTGACGAGTTTGATTATGAATGGCCATGTCACC 180
QY 832 CGCTACACAAAGGAGATGAGGCTATGCCGCCCAAGAGAGAACCAAACTGCTGAA 891
Db 181 CGCTGCACAAAGGAGATGAGGCTACGCGCTCCCAAGAGAGAACCAAACTGCGGNA 240
QY 892 AGGCGCAAGCGTGTCTGAGGAGCACATCTACAGGAAATGATGACATGTGCTTATTC 951
Db 241 AGGCGTAAAGAGCTGAGGAGCACATCTACAGAGAAATCATGACATGTGCTTATTC 300
QY 952 TGCACATGGCTGCCACACAGAGATGCCAGATCCAGTTCAGTTTTCGAGATCTCTTT 1011
Db 301 CGCACATGGCTGCCACCCCGGAGCGGACGCAAGATCCAGTTTACITTCGAGATCTCTTT 360
QY 1012 GACAGATAGCTTCGTATTTTCAGATAAAGTAGTGGGCATTTCTATCGTGGCCAGGAAACAT 1071
Db 361 GACAGATAGCTTCGTATTTTCAGATAAAGTAGTGGGCATTTCTATCGGCGCCAGGAAACAC 420
QY 1072 GACTGCTAGACTTTGAAGAGAGATGCTATGCGAGGCGGAGATGACCATGCTGTGATT 1131
Db 421 GACTGCTAGACTTTGAGGAGAGATGCTATGCGAAGGCGGAGATGACCACGTTGTGATC 480
QY 1132 ACGTACTCTCAAGTGA 1146
Db 481 ACTCTGCTAAAGTGA 495

RESULT 13
BF549525 553 bp mRNA linear EST 12-DEC-2000
LOCUS UI-R-C1-li-h-06-0-UI-r1 UI-R-C1 Rattus norvegicus cDNA clone
DEFINITION UI-R-C1-li-h-06-0-UI 5', mRNA sequence.
ACCESSION BF549525
VERSION BF549525.1 GI:11659213
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 553)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL 97044477
MEDLINE 8889548
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone distribution:
This clone will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LNL (info@image.llnl.gov). IMAGE ID= 1775884
Seq primer: M13 Forward.
Location/Qualifiers

```

source

```

1..553
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C1-li-h-06-0-UI"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-C1"

```

```

/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C1
library is a subtracted library derived from the UI-R-C0
library, which is a subtracted library derived from the
UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library
consisted of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung,
brain, liver, kidney, heart, spleen, ovary, and muscle.
The UI-R-E1 library consisted of a mixture of
individually tagged normalized libraries constructed from
8, 12 and 18-day embryo. The tag is a string of 3-5
nucleotides present between the Not I site and the
oligo-dr track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-C1) was constructed as follows: PCR
amplified cDNA inserts from UI-R-C0 clones from which 3'
ESTs had been derived was used as a driver in a
hybridization with the UI-R-C0 library in the form of
single-stranded circles. The remaining single-stranded
circles (subtracted library) was purified by
hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the UI-R-C1
library. This procedure has been previously described
(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
1996)."

```

ORIGIN

```

Query Match      36.4%; Score 416.6; DB 10; Length 553;
Best Local Similarity 85.3%; Pred. No. 2.3e-95;
Matches 464; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 543 GCCACATGAGGAGTGCACGCTAGACACAGAGACAGCGGCTATGGAGAGAGCTGA 602
Db 7 GCCCAAGTGGAGAGTGCACATAGACACAGAGACAGTGGCTACGAGGGGACATGA 66
QY 603 GGAGAGGCCCGAGCAGGATGGAGTGCAGTGGCTGTGCTCAGGATCAAGCCGCCCTTGCC 662
Db 67 GGAGAGGCCCTGAGCAAGATGTAGCGCAGTGGCTGTGCTCAGGATTAACGCCCTTGCA 126
QY 663 CTCCAGGTAAACAGATTTCAGAGAACTCAACTGCAAGCCCAACAGAAATATAGCCC 722
Db 127 CTCCAGGCAAAACAGATCTCAGAGACACTCAACTGTAAGGCCCATCGAAATACAGCCA 186
QY 723 AGTGGCAACTTGAAGGGAGATGGCAGCAGTGGGCTGTGCTCAGGATCAACATCCAGAA 782
Db 187 AGTGACAACTTGAAGAGCAGTGGCAGCAGTGGGCGATGAACATACATAGTACAGAA 246
QY 783 GCTCAATCCTTTCACTGAGAGATTGATTACAGCTGGCCATGTCACCGCCCTACACAA 842
Db 247 GCTCAACCCCTTCAGTGATGAATTTGACTATGACCTAGCCATGTCACCTCGACTCCACA 306
QY 843 AGGAGATGAGGCTATGGCCGCCCAAGAGAGACCAAACTGCTGAAGAGGCCCAAGCG 902
Db 307 GGGAGACGAGGATATGGCCGCCCAAGAGAGAGACAGCTGAAGAGGCCCAAGAG 366
QY 903 TGCTGAGGAGCACATCTACAGGAAATGATGACATGTGCTTCAATTCGACAAATGCG 962
Db 367 ACCCGAGGAGCACATCTATCGGAAATTTATGGAATTTGCTTTGTTATCCGCACAAATGGC 426
QY 963 TCGCCACAGACAGATGGCAAGATCCAGGTACITTTTGAGATCTCTTTTGACAGATAGT 1022
Db 427 TCGCCATAGACAGATGGCAAGATCCAGGTACTTTTCGAGAGAACTCTTTTGTAGATAGT 486

```


QY 1023 TCGTATTTCAGATAAGTAGGGGCAATCTCATCGTGCAGGAAACATGACTGGTGA 1082
 |||||
 Db 487 TCGAATTTTCAGATAAAGTGGTGGGCATCTCATCGTGCAGGAAACACAGACTGGTGA 546
 |||||
 QY 1083 CTTT 1086
 |||||
 Db 547 CTTT 550

RESULT 14
 BX671003
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Sus scrofa (pig)
 Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 Bonnet A., Tosser-Klopp G., Benne, F., Cabau, C., Villegier, S.,
 Soares, M., Bonaldo, F. and Hately, F.
 A Pig Normalised Multi-Tissue cDNA Library
 Unpublished (2003)
 Contact: Tosser-Klopp G
 Genetique Animale
 Institut National de la Recherche Agronomique
 Chemin de Borde-Rouge - Auzeville BP27, 31326 Castanet-Tolosan
 cedex, FRANCE
 Tel: 33 (0) 5.61.28.51.14
 Fax: 33 (0) 5.61.28.53.08
 Email: tosser@toulouse.inra.fr
 Clone distribution: AGENAE Resource centre, Francois PIUMI,
 Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et Etude du
 genome (LREG), Domaine de Vilvert, 78352, Jouy-en-Josas cedex,
 FRANCE. +33 (0) 1.34.65.28.02. +33 (0) 1.34.65.22.73
 Sequence cleaned of vector, adaptor and repetitions. Contact us
 at signenasupport@jouy.inra.fr to obtain the chromatogram of this
 sequence.
 Plate: 0035 row: h column: 22.
 Location/Qualifiers
 1. 630
 /organism="Sus scrofa"
 /mol_type="mRNA"
 /db_xref="taxon:9823"
 /clone="scac00351.h.22"
 /tissue_type="mixed"
 /clone_lib="Sus Scrofa library (scac)"
 /note="Vector: pT73b-pac vector; tissues: adipose tissue,
 brain, kidney, liver, muscle, ovary, testis, heart,
 hypothalamus, pancreas, skin, spleen, thymus, placenta,
 pituitary gland, seminal vesicle, small intestine,
 uterus, adrenals, bulbo urethral gland, cerebral trunk,
 epididymis, female gonad, gall-bladder, hippocampus,
 large intestine, male gonad, melanocytes, stomach, udder"

FEATURES
 source
 1. 630
 /organism="Sus scrofa"
 /mol_type="mRNA"
 /db_xref="taxon:9823"
 /clone="scac00351.h.22"
 /tissue_type="mixed"
 /clone_lib="Sus Scrofa library (scac)"
 /note="Vector: pT73b-pac vector; tissues: adipose tissue,
 brain, kidney, liver, muscle, ovary, testis, heart,
 hypothalamus, pancreas, skin, spleen, thymus, placenta,
 pituitary gland, seminal vesicle, small intestine,
 uterus, adrenals, bulbo urethral gland, cerebral trunk,
 epididymis, female gonad, gall-bladder, hippocampus,
 large intestine, male gonad, melanocytes, stomach, udder"

Query Match 35.0%; Score 401; DB 13; Length 630;
 Best Local Similarity 82.2%; Pred. No. 2.5e-91;
 Matches 488; Conservative 0; Mismatches 96; Indels 10; Gaps 2;

QY 1 ATGGGTCGGGCGAAGAAAGAGCGGGAGGCGCCAGAGCGCCCTCC-GGAAGAT 59
 |||||
 Db 37 ATGGCGCGGAGAAAGGAGAGAGGGAGGACCTGCGAAGTGCCTCCNGGAGGT 96
 |||||
 QY 60 AGCGACAGCCCTCGTTCATCAGCTTGGCCCGAGGTGGCAGCAGTGGCGAATGAGAA 119
 |||||
 Db 97 AGCGACAGCCCTCGTTCATCAGCTTGGCCCGGAGGTGGCAGCAGTGGCGAATGAGAA 156
 |||||
 QY 120 CAGCATCAGGAGGCGCAGAGCCCTACAGCTGGCTGCGGAGGAGGCCAGGACTCACC 179
 |||||

Db 157 TAGCACCAGGAGGCGCCAGGAGCCTACAGGTGGATGCCAGGAGGGCGAGGAGTCAGA 216
 QY 180 TCAAGCTCTTAAACCAATCACACCCCTACTTTCACACAGAAAGCTCAGAGTGGCCCAA 239
 |||||
 Db 217 CCAACCTTCTGGACCAAGTGTATACACCTTACACCCACAGAAAGTCCAGAGTGGCCCAA 276
 |||||
 QY 240 GTCCGCACCCCGCTGCCAGAGAGGACATGGAGATCGACAAAGCTCAGAGAAAGCCCTTGA 299
 |||||
 Db 277 ATCTCCCTCCCAAAGCCAGGGGGATATGGAGTGTGACAAAGCTCAGAGGGAGCCACCGA 336
 |||||
 QY 300 GGTTCCTCATCAATCAAAAGAAAGAGGTGTCCAAACGGTGTGTGTCAGCAAGACTTACGAGAG 359
 |||||
 Db 337 AGTCTCTCTATCAAAAGGAAAGAGGTGACCAAGACCTATGTGACAAAGCTTACGAGAG 396
 |||||
 QY 360 AGGAGGGGACGTGAGCCACCTCAGCCACAGGATACAGAGGGATGCTGTGTCTTGAACC 419
 |||||
 Db 397 AGGAGTGTATGTAAGCCACCTCAGCATAGATACAGAGGATGTGTGATGAGCTTGAACC 456
 |||||
 QY 420 TGGGACGACAGAGATGACATTCAGAGAAATCTTCACAGCCACGGCTCCCAACGGGAG 479
 |||||
 Db 457 CGAGCAGCCAGAGAGTGACATTCAGAGACTCTCTCCGACCCATGGCTCCCAACGGGAG 516
 |||||
 QY 480 GAGAAATGTGCCAACCTGTGTCTGAGTAAACCAAGGCTGGAGAGTGAT----- 530
 |||||
 Db 517 AGAAATGTGCCAACCTGTGTCTGAGTAAACCAAGGCTGGAGAGTGATGGAACAGGA 576
 |||||
 QY 531 GGAGCAGGAGGAGCCACATGAGGAGTACAGCTAGACACAGAGGACAGCGG 584
 |||||
 Db 577 GGAACAGGAGGAGCTCAATGACAGGATGATAGCATAGACACAGAGAGCAGCGG 630
 |||||

RESULT 15
 BX511434
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Heil, O., Ebert, L., Neubert, P., Peters, M., Radelof, U., Schneider, D.
 and Korn, B.
 Mouse Unigeneset - RZPD2
 Unpublished (2003)
 Contact: Ina Rolfs
 RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
 Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
 RZPD: IWAGP998J06708.
 RZPDLIB: I.M.A.G.E. cDNA Clone Collection:
 Mouse Unigeneset - RZPD2 (RZPDLIB No.981)
 http://www.rzpd.de/Clonecards/cgi-bin/showlib.pl.cgi/response?libNo=981 Contact: Ina Rolfs
 RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
 Heubnerweg 6, D-14059 Berlin, Germany
 Tel: +49 30 32639 101
 Fax: +49 30 32639 111
 www.rzpd.de

This clone is available royalty-free from RZPD;
 contact RZPD (clone@rzpd.de) for further information. Seq primer:
 T7, Primer sequence: TAATACGACTCACTATAGG.

FEATURES
 source
 1. 528
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGP998J06708 ; IMAGE:314213"
 /dev_stage="19.5 dpc total fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares mouse p3NMf19.5"

/note="Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dN) primer [5', TGTTACCACTGATGAGCGACGCCCATTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Faúma Bonaldi. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

ORIGIN

Query Match.	34.3%;	Score 392.6;	DB 13;	Length 528;
Best Local Similarity	84.1%;	Pred. No. 3.1e-84;		
Matches 443;	Conservative 0;	Mismatches 84;	Indels 0;	Gaps 0;
QY	406	GGTGTGCTTGAACCTCGGCGAGCCAGAGAAATGCACATTGCACAGAAATCCCTCCACAGCCACGGC	465	
DB	2	GGCGTGCTGAAGCTATTACGCCAGAGAAATGCACATTGCACAGAAATCCCTTTAGTCACGAC	61	
QY	466	TCGCCAACGGGAGGAGAAATGTGCCAAACCTGTGTCTGAGCTTAACCAAGGGCTGGAGA	525	
DB	62	TCGCCAACCGGAGAGAAATGCACCACTGTGTCTGAGCTGACCAAGGCTGGAAA	121	
QY	526	GTGATGAGCAGGAGAGGCCACATGAGAGAGTGCACGCTAGACACAGAGGACAGCCGC	585	
DB	122	GTGATGAAACAGGAAGAGCCACCTGTGGAAGAGTGCACGCTAGACACAGAGGACAGTGGC	181	
QY	586	TATGGAGGAGGCTGAGGAGAGCCCGAGCAGCATGGAGTGCAGGTGGCTGTGGTCAGG	645	
DB	182	TACGGAGGGGATATGGAGGAGGCGCTTGAGCAGATGCAGCGCTGTGGCTCTCCGACGG	241	
QY	646	ATCAAGGGCCCTTTGCCCTCCCGAGTAAACAGATTTACAGAGAAATCAACTGCAAAAGCC	705	
DB	242	ATCAACGCCCCCTTGTCTCTCCAGCAACAGGTAATCTGAGACACTCAACTGTAAAGCC	301	
QY	706	CAACAGAAATATAGCCAGTGGGCAACTTTGAAAGGAGATGGCAGCAGTGGGCTGATGAA	765	
DB	302	CATCGGAAATACAGCCAAGTGGCAACTTTGAAAGGAGGTGGCAGCAGTGGGGCCGATGAA	361	
QY	766	CACATACAATCCAGAGAGCTCAATCCCTTTTCAGTCAAGAGTTTGATTTACGAGCTGGCCATG	825	
DB	362	CACGTTCAGTCCCAGAAAGCTCAATCCCTTTTCAGTACGAAATTTGACTATGATCTTASCCATG	421	
QY	826	TCCACCCGCTTACCAAGAGGAGATGAGGGCTATGGCCGCCCCCAAGAGAGAACCAAAACT	885	
DB	422	TCCACTCGCTCCACAGGGAGACGAGGGCTATGGCCGCCCCCAAGAGGGAAGCAAGACA	481	
QY	886	GCTGAAAGGGCCAAAGCGGTGTGAGAGACATCTTACAGGGAAATGAT	932	
DB	482	GCTGAAAGGGCCAAAGCGAGCGGAGAGACATCTTATCGGGAATAT	528	

Search completed: May 5, 2004, 03:03:18
Job time : 2878 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM nucleic - protein search, using frame_plus_n2p model

Run on: May 4, 2004, 08:51:39 ; Search time 60.5 Seconds

(Without alignments)

10501.133 Million cell updates/sec

Title: US-10-644-659A-1

Perfect score: 2058

Sequence: 1 atggctccggggaagga.....tgattacgctactcaagtga 1146

Scoring table:

BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 1138120 seqs, 277189581 residues

Total number of hits satisfying chosen parameters: 2276240

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE/frame+ n2p.spool p/US10644659/runat 04052004 084903 28592/app query.fasta_1.1287
-Q/cgn2 1/USPro model p/US10644659/runat 04052004 084903 28592/app query.fasta_1.1287
-DB=published Applications AA -QFMT=fastn -SURFI=n2p.rapb -MINMATCH=0.1
-LOPCH=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10644659 @CGN 1.13 @runat 04052004 084903 28592
-NCPU=6 -ICPU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCKS=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=10 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
-----	-----	-----	-----	-----	-----

1	2029	98.6	381	15	US-10-104-047-3169	Sequence 3169, Ap
2	647	31.4	121	9	US-09-864-761-42191	Sequence 42191, A
3	190	9.2	37	9	US-09-864-761-42192	Sequence 42192, A
4	142	6.9	19695	15	US-10-084-846A-3	Sequence 3, Appli
5	134.5	6.5	19662	15	US-10-084-846A-6	Sequence 6, Appli
6	133	6.5	19608	15	US-10-084-846A-8	Sequence 8, Appli
7	130.5	6.3	1739	15	US-10-341-434-61	Sequence 61, Appli
8	129.5	6.3	19723	15	US-10-084-846A-5	Sequence 5, Appli
9	126.5	6.1	93	12	US-10-425-114-63655	Sequence 63655, A
10	126.5	6.1	19652	15	US-10-084-846A-7	Sequence 7, Appli
11	125.5	6.1	645	12	US-10-425-114-4380	Sequence 4380, A
12	124.5	6.0	339	12	US-10-425-114-45434	Sequence 45434, A
13	123.5	6.0	422	12	US-10-425-114-70922	Sequence 70922, A
14	123.5	6.0	460	12	US-10-425-114-71620	Sequence 71620, A
15	122	5.9	293	14	US-10-029-386-33391	Sequence 33391, A
16	121.5	5.9	506	14	US-10-156-761-13648	Sequence 12648, A
17	121	5.9	284	14	US-10-029-386-33066	Sequence 33066, A
18	121	5.9	488	14	US-10-365-742-20	Sequence 20, Appli
19	121	5.9	1182	14	US-10-024-368-6	Sequence 6, Appli
20	120.5	5.9	1019	12	US-10-276-774-2298	Sequence 2298, Ap
21	120.5	5.9	2803	12	US-10-415-187-5	Sequence 5, Appli
22	120	5.8	485	14	US-10-091-752A-14	Sequence 14, Appli
23	120	5.8	485	14	US-10-136-547-14	Sequence 14, Appli
24	119.5	5.8	375	14	US-10-156-761-14622	Sequence 14622, A
25	119.5	5.8	442	12	US-10-425-114-40452	Sequence 40452, A
26	119	5.7	19695	15	US-10-084-846A-3	Sequence 3, Appli
27	118.5	5.8	311	12	US-10-425-114-61343	Sequence 61343, A
28	117.5	5.6	524	15	US-10-289-762-52	Sequence 52, Appli
29	116.5	5.7	91	12	US-10-424-599-237720	Sequence 237720, A
30	116.5	5.7	336	12	US-10-425-114-56601	Sequence 56601, A
31	116.5	5.7	685	15	US-10-104-047-2916	Sequence 2916, Ap
32	115.5	5.6	656	15	US-10-108-260A-2611	Sequence 2611, Ap
33	115.5	5.6	19725	15	US-10-084-846A-4	Sequence 4, Appli
34	115	5.6	629	15	US-10-384-919-2	Sequence 2, Appli
35	114.5	5.6	300	12	US-10-425-114-64837	Sequence 64837, A
36	114.5	5.6	336	12	US-10-425-114-68820	Sequence 68820, A
37	114	5.5	383	12	US-10-425-114-53657	Sequence 53657, A
38	113.5	5.5	324	12	US-10-425-114-53603	Sequence 53603, A
39	113.5	5.5	362	12	US-10-425-114-59602	Sequence 59602, A
40	113.5	5.5	367	12	US-10-425-114-68177	Sequence 68177, A
41	113.5	5.5	367	12	US-10-425-114-65068	Sequence 65068, A
42	113	5.5	324	12	US-10-425-114-47752	Sequence 47752, A
43	113	5.5	1593	11	US-09-981-151A-30	Sequence 30, Appli
44	113	5.5	1593	15	US-10-295-027-1317	Sequence 1317, Ap
45	113	5.5	2038	15	US-10-433-794-18	Sequence 18, Appli

ALIGNMENTS

RESULT 1

US-10-104-047-3169
; Sequence 3169, Application US/10104047
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1el full length cdNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3169
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3169

Alignment Scores:	1.11e-158	Length:	381
Pred. No.:	2029.00	Matches:	381
Score:	100.00%	Conservative:	0

```
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.59% Indels: 0
DB: 15 Gaps: 0
```

US-10-644-659A-1 (1-1146) x US-10-104-047-3169 (1-381)

Qy	1	ATGCTCCGGCGGCGAAAGAAAGAAAGCGGGAGGGCGCCAGCGCCCTCCGGAGAGATA	60
Db	1	MetAlaProGlyGluLysGluSerGlyGluGlyProAlaLysSerAlaLeuArgLysIle	20
Qy	61	CGACAGCCACCCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGCGGAATGAGAAC	120
Db	21	ArgThrAlaThrLeuValIleSerLeuAlaArgGlyTrpGlnGlnTrpAlaAsnGluAsn	40
Qy	121	AGCATCAGCAGGCCCGCCAGGAGCTTACAGGTGGCTGGCTCCGGAGGGAGCCACGACTCACCT	180
Db	41	SerIleArgGlnAlaGlnGluProThrGlyTrpLeuProGlyGlyThrGlnAepSerPro	60
Qy	181	CAAGCTCCTAAACCAATCACACCCCTPACTTCACACCAGAAAGCTCAGAGTGCCCAAAG	240
Db	61	GlnAlaProLysProIleThrProProThrSerHisGlnLysAlaGlnSerAlaProLys	80
Qy	241	TCGCCACCCCGCTCCGCAAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCTCGAG	300
Db	81	SerProProArgLeuProGluGlyHisGlyAaspGlyGlnSerSerGluLysAlaProGlu	100
Qy	301	GTTTCTCACATCAAAAAGAAAGAGTCTCAAACAGGTGCTCAGCAAGACTTACGAGAGA	360
Db	101	ValSerHisIleLysLysLysGluValSerLysThrValSerLysThrTrpGluArg	120
Qy	361	GGAGGGAGCTGAGCCACCTCAGCCA CAGGTACGAGAGGATGCTGGTGTGCTTGAACCT	420
Db	121	GlyGlyAaspValSerHisLeuSerHisArgTyrGluAArgAaspAlaGlyValLeuGluPro	140
Qy	421	GGGCAGCCAGAGATGACATTGCACAGAAATCTCTCCACAGCCACGCTCCCCAACCGCGAGG	480
Db	141	GlyGlnProGluAsnAspIleAaspArgIleLeuHisSerHisGlySerProThrArgArg	160
Qy	481	AGAAAAATGTGCAACCTGGTCTCGAGCTAACCAAGGGCTGGAGATGATGGAGCAGGAG	540
Db	161	ArgLysCysAlaAsnLeuValSerGluLeuThrLysGlyTrpArgValMetGluGlnGlu	180
Qy	541	GAGCCACATGGAGCAGTGACAGCGTATAGACACAGAGCAGCAGCGCTATGAGGAGAGGCT	600
Db	181	GluProThrTrpArgSerAaspSerValAaspThrGluAaspSerGlyTyrGlyGlyGluAla	200
Qy	601	GAGGAGAGGCCCGAGCAGGATGGAGTGGAGTGGCTGTGTCAGGATCAAGCGCCCTTG	660
Db	201	GluGluArgProGluGlnAaspGlyValGlnValAlaValValArgIleLysArgProLeu	220
Qy	661	CCCTCCCAAGTAAACAGATTTCACAGAGAACTCAACTGCAAGCCCAACAGAAATATAGC	720
Db	221	ProSerGlnValAsnArgPheThrGlnLysLeuAsnCysLysAlaGlnGlnLysTyrSer	240
Qy	721	CCAGTGGGCAACTTGAAGGGAGATGGCAGCAGTGGGCTGATGAACACATACAAATCCCAG	780
Db	241	ProValGlyAsnLeuLysGlyArgTrpGlnGlnTrpAlaAaspGluHisIleGlnSerGln	260
Qy	781	AAGCTCAATCCTTTCAGTGAAGAGTTTGATTACAGCTGGCCATGTCCACCCGCCACAC	840
Db	261	LysLeuAsnProPheSerGluGluPheAspTyrGlnLeuAlaMetSerThrArgLeuHis	280
Qy	841	AAAGGAGATGAGGCTATGGCCGCCCCCAAGAGAAAGCAAACTGTGAAAGGGCCCAAG	900
Db	281	LysGlyAaspGluGlyTyrGlyArgProLysGluGlyThrLysThrAlaGluAArgAlaLys	300
Qy	901	CGTCTGAGGAGCACATCTACAGGGAATGATGACATGTGCTTCATTCTGCACCAATG	960
Db	301	ArgAlaGluGluHisIleTyrArgGluMetMetAaspMetCysPheIleIleCysThrMet	320
Qy	961	GCTCGCCACAGACGAGATGGCAAGATCCAGTTACTTTTGAGATCTCTTTGACAGATAC	1020
Db	321	AlaArgHisArgAArgAaspGlyLVleGlnValThrPheGlyAaspLeuPheAaspArgTyr	340

RESULT 2

US-09-864-761-42191
; Sequence 42191. Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: aeomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 42191
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC023344.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2

```
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
; OTHER INFORMATION: EST HUMAN HIT: BF083249.1, EVALUE 8.50e-01
; OTHER INFORMATION: SWISSPROT HIT: P03189, EVALUE 8.80e+00
US-09-864-761-42191

Alignment Scores:
Pred. No.:      8.54e-45      Length:      121
Score:          647.00      Matches:      121
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      31.44%      Indels:      0
DB:              9      Gaps:      0

US-10-644-659A-1 (1-1146) x US-09-864-761-42191 (1-121)
QY 262 GGACATGGAGATGGACAAAGCTCAGAGAAAGCCCTGAGGTTTCTCACATCAAAAGAAA 321
Db 1 GlyHisGlyAspGlyGlnSerSerGlyLysAlaProGluValSerHisIleIleLysLys 20
QY 322 GAGGTGTCGAAACGGTGTCTCAGACACTTACGAGAGAGGGGGGACGTGACCTC 381
Db 21 GluValSerLysThrValValSerLysThrThrGluArgGlyGlyAspValSerHisLeu 40
QY 382 AGCCACAGGTACAGAGGGATGCTGCTGCTGCTTGAACCTGGGCGAGCCAGAGATGACATT 441
Db 41 SerHisArgTyrgluargaspAlaGlyValLeuGluProGlyGlnProGluAsnAspIle 60
QY 442 GACAGAAATCTCCACAGCCACGGCTCCCAACCGGAGAGAGAAATGTGCCAACCTGGTG 501
Db 61 AspArgIleLeuHisSerHisGlySerProThrArgArgLysCysAlaAsnLeuVal 80
QY 502 TCTGAGCTAACAGGCTCGAGAGTGTGAGCAGCAGAGAGCCACATGGAGGAGTGAC 561
Db 81 SerGluLeuThrLysGlyIleArgValMetGluGlnGluProThrIleArgSerAsp 100
QY 562 AGCGTAGACACAGAGGACCGGCTATGGAGGAGGCTGAGAGAGGCGCCGAGCAGGAT 621
Db 101 SerValAspThrGluAspSerGlyTyrglyGlyGluAlaGluArgProGluGlnAsp 120
QY 622 GGA 624
Db 121 Gly 121

RESULT 3
US-09-864-761-42192
; Sequence 42192, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30

; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
; OTHER INFORMATION: EST HUMAN HIT: P03189, EVALUE 8.80e+00
; OTHER INFORMATION: SWISSPROT HIT: P03189, EVALUE 8.80e+00
US-09-864-761-42191

Alignment Scores:
Pred. No.:      8.54e-45      Length:      121
Score:          647.00      Matches:      121
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      31.44%      Indels:      0
DB:              9      Gaps:      0

US-10-644-659A-1 (1-1146) x US-09-864-761-42191 (1-121)
QY 262 GGACATGGAGATGGACAAAGCTCAGAGAAAGCCCTGAGGTTTCTCACATCAAAAGAAA 321
Db 1 GlyHisGlyAspGlyGlnSerSerGlyLysAlaProGluValSerHisIleIleLysLys 20
QY 322 GAGGTGTCGAAACGGTGTCTCAGACACTTACGAGAGAGGGGGGACGTGACCTC 381
Db 21 GluValSerLysThrValValSerLysThrThrGluArgGlyGlyAspValSerHisLeu 40
QY 382 AGCCACAGGTACAGAGGGATGCTGCTGCTGCTTGAACCTGGGCGAGCCAGAGATGACATT 441
Db 41 SerHisArgTyrgluargaspAlaGlyValLeuGluProGlyGlnProGluAsnAspIle 60
QY 442 GACAGAAATCTCCACAGCCACGGCTCCCAACCGGAGAGAGAAATGTGCCAACCTGGTG 501
Db 61 AspArgIleLeuHisSerHisGlySerProThrArgArgLysCysAlaAsnLeuVal 80
QY 502 TCTGAGCTAACAGGCTCGAGAGTGTGAGCAGCAGAGAGCCACATGGAGGAGTGAC 561
Db 81 SerGluLeuThrLysGlyIleArgValMetGluGlnGluProThrIleArgSerAsp 100
QY 562 AGCGTAGACACAGAGGACCGGCTATGGAGGAGGCTGAGAGAGGCGCCGAGCAGGAT 621
Db 101 SerValAspThrGluAspSerGlyTyrglyGlyGluAlaGluArgProGluGlnAsp 120
QY 622 GGA 624
Db 121 Gly 121

RESULT 4
US-10-084-846A-3
; Sequence 3, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUEHLEWEG, AGNES
; APPLICANT: TREPZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
```


DB: 15 Gaps: 10

US-10-644-659A-1 (1-1146) x US-10-341-434-61 (1-1739)

QY 109 GCGAATGACAGACAGCATCAGGAGCCAGGAGCTACAGGTGCTCCGGAGGAGCC 168

Db 17 SerAsnAlaSerSerHisSerAlaSerGluGluAlaSerGlySerAspSerGlySerGln 36

QY 169 CAGGACTCACCTCAGCTCCTAAACCAATCACACCCCTACTTCAAC--CAGAAAGCT 225

Db 37 SerGluSerGluGlnGlySer-----AppProGlySerGlyHisGlySerGluSer 53

QY 226 CAGAGTCCCAAGAGTCGCGCCAGGAGCATGAGATGAGATGAGCAAGCTCA 285

Db 54 AsnSerSerSerGluSerGluSerGluSerGluSerGluSerGluSerGluSer 73

QY 286 GAGAAAGCCCTCAGCTTCTCATCAAAAGAGAGGTGTCCTCAAAAGCTGGTCAGC 345

Db 74 LysSerGlnProValLeuProGluAlaLysProAlaSerLysLys-----90

QY 346 AAGACTTACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 405

Db 91 -----GluArgileAlaAspValLysLysMetTTPGluGluTyrProAspValTyr 107

QY 406 GTGTGCTTGAACCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 453

Db 108 GlyValArgArgSerSerSerArgGlnGluProSerArgSerPheAsnLysGluGlu 127

QY 454 -----CACAGCCAGGCTCCCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 501

Db 128 AlaserSerGlySerGluSerGlySerProLysArgArg-----140

QY 502 TCTGAGCTAACAGGCTGAGAGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 561

Db 141 -----GlyGlnArgGlnLeuLysLysGlnGluLysLysGlnGluPro 155

QY 562 AGCTAGACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 621

Db 156 SerGluAspGluGlnGlnGlnGlyThrSerAlaGluSerGlu-----ProGluGln---172

QY 622 GGAGTGCAGGTGCTGTGTCAGATCAAGCGCCCTTCCCTCCAGGTAAACAGATT 681

Db 173 -----LysLysValLysAlaArgProValPro-----ArgArgThr 185

QY 682 ACAGAGAACTCAACTGCAAGGCCCAACAGAAATATAGCCAGTGGGCAACTTGAAGGG 741

Db 186 ValProLysProArgValLysLysGlnProLys-----ThrGlnArgGlyLysArg 202

QY 742 AGATGGCAGCAGTGGCTGATGAA-----765

Db 203 LysLysGlnAspSerSerSerAspGluAspAspAspAspGluAlaProLysArgGlnThr 222

QY 766 CATATCAATCCAGAGAGCTCAATCTTTCAGTGAAGAGTTTATACGAGCTGGCCATG 825

Db 223 ArgArgArgAlaAlaLysAsnValSerTyrLysGluAspAspPheGluThrAspSer 242

QY 826 TCCACCGCCTCACAAAGAGAGATGAGGCTATGGCGCCGCCCAAGAGAGGAGGAGGAG 885

Db 243 AspAspLeuileGluMetThrGlyGluGlyValAspGluGlnGlnAspAsnSerGluThr 262

QY 886 GCTGAAAGG 894

Db 263 IleGluLys 265

RESULT 8

US-10-084-846A-5

; Sequence 5, Application US/10084846A

; Publication No. US20040006026A1

; GENERAL INFORMATION:

; APPLICANT: WEITNAUER, GABRIELE

; APPLICANT: MUHLENWEG, AGNES

; APPLICANT: TREPZER, AXEL

; APPLICANT: BECHTHOLD, ANDREAS

QY 500 TGTCTGAGCTAACCAAGG-----517

Db 5380 ArgArgCysTTPProArgTTPSerArgAsnProArgProCysAlaThrSerTTPArgThr 5399

QY 518 -----GCTGGAGAGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 565

Db 5400 AlaSerThrSerArgAlaTyrGlyTTPSerArgThrSerPro-----ThrSer 5415

QY 566 TAGACACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 622

Db 5416 ProThrAlaSerThrAlaProAlaThrThrSerProAlaThrProProAlaSerThr 5435

QY 623 GAGTGC-----AGGTGGCTGTGGTCAGGATCAAGGCGC-----CCT 658

Db 5436 ArgCysCysProProAlaSerThrTTPArgSerThrAlaAlaCysTTPProArgArgPro 5455

QY 659 TGCCCTCCAGTAAACAGATTACAGAGA-----AACTCA 694

Db 5456 CysSerProArgThrArgThrSerArgArgArgProGlyArgSerThrSerArgSer 5475

QY 695 ACTGCAAGGCCCAACAGAAATATACCCAGTGGGCAACTTGAAGGGAGAGTGGCAGCAGT 754

Db 5476 ThrAlaThrProThrSerGlyCysSerProTTPCysPro-----5488

QY 755 GGGCTGATGAACACATACATCCCAAGAGCTCAATCTTTCAGTGAAGAGTTTGAATACG 814

Db 5489 -----AlaCysThrSerAsnArgProAlaArgProProThrSer 5501

QY 815 AGCTGGCCATGTCACCCGCCCTACACAAAGGAGATGAGGCTATGGCGCCCAAGAGAG 874

Db 5502 SerTTPProThrArgPro-----ArgSerAlaProArg---5512

QY 875 GAACCAAACTGCTGAAAGGGCCAGCGTGTGAGGAGCACATCTACAGGGAATGATGG 934

Db 5513 -----ProSerThrSerArgSerThrAlaArg-----TTP 5522

QY 935 ACATGTGCTTCATTTCTCCAAATGG-----CTCGCCACAGACGAGATGGCAAGATCC 988

Db 5523 ProSerProAsnTTPSerArgAspTTPProThrSerThrProSerProAlaArgThr 5542

QY 989 AGG 991

Db 5543 Arg 5543

RESULT 7

US-10-341-434-61

; Sequence 61, Application US/10341434

; Publication No. US20030215835A1

; GENERAL INFORMATION:

; APPLICANT: Origene Technologies

; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes

; FILE REFERENCE: 90 204 205 R1

; CURRENT APPLICATION NUMBER: US/10/341,434

; CURRENT FILING DATE: 2003-07-18

; PRIOR APPLICATION NUMBER: US 60/348,164

; PRIOR FILING DATE: 2002-01-15

; PRIOR APPLICATION NUMBER: US 60/348,119

; PRIOR FILING DATE: 2002-01-15

; NUMBER OF SEQ ID NOS: 238

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 61

; LENGTH: 1739

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-341-434-61

Alignment Scores:

Pred. No.: 0.0642 Length: 1739

Score: 130.50 Matches: 63

Percent Similarity: 37.10% Conservativeness: 42

Best Local Similarity: 22.26% Mismatches: 123

Query Match: 6.34% Indels: 55

```

/ TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
/ FILE REFERENCE: 1974-005
/ CURRENT APPLICATION NUMBER: US/10/084,846A
/ CURRENT FILING DATE: 2003-02-25
/ PRIOR APPLICATION NUMBER: PCT/EP01/09815
/ PRIOR FILING DATE: 2001-08-24
/ PRIOR APPLICATION NUMBER: DE 101 09 166.4
/ PRIOR FILING DATE: 2001-02-25
/ NUMBER OF SEQ ID NOS: 120
/ SOFTWARE: PatentIn Ver. 3.2
/ SEQ ID NO 5
/ LENGTH: 19723
/ TYPE: PRT
/ ORGANISM: Streptomyces viridochromogenes
/ FEATURE:
/ OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.
/ OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.
US-10-084-846A-5

Alignment Scores:
Pred. No.: 0.137 Length: 19723
Score: 129.50 Matches: 99
Percent Similarity: 28.82% Conservatives: 33
Best Local Similarity: 21.62% Mismatches: 147
Query Match: 6.29% Indels: 179
DB: 15 Gaps: 20

US-10-644-659A-1 (1-1146) x US-10-084-846A-5 (1-19723)
QY 12 CGAAAGGAAGCGGGAGGCGCCAGCAGAGCCCTCCGGAAGATACGACAGCCAC 71
DB 48 ArgArgGlyProArg---ProProGlyHisArgAspProProHisProArgHisGlyHis 66
QY 72 ---CCTGTCATCATGCTTGGC----- 89
DB 67 ArgGlnArgAlaValHisArgArgProGlyProGlnHisGlyAlaGlyProProArg 86
QY 90 -----CCGAGGTGGCAGCATGGCGGCGAATGAGACACATCATGAGGCCAGGCCCCAGGA 140
DB 87 ThrProTrpProArgArgGluAla-----ArgLeuGlnGlnGlyGlnGlyGlyArgGly 104
QY 141 GCCTACAGGTGGCTGCGGGAGGAGGCCAGGAGCTCACCTCAAGCTCCTAAACCA----- 195
DB 105 AlaGluLeuGlyAlaLeuGly-----Pro-ProArgProGluValProArgProCysArg 122
QY 196 -----ATCACACCCCTACTTCACACAGAAAGCTCAGAGTGCCCCCAAGTCGCCA-- 246
DB 122 GASPAlaValLeuProGlyAlaValHisArgArgArgGlyGluProGlnValProGln 142
QY 247 ----CCCCGCTGCGAAGACATGAGATGACAAAGCTCAGAGAAAGCCCTCAGAGT 302
DB 142 nValProAlaProHisArgGlyHisGlnGlyGlyAspArgProGlyProAlaAlaGlu-- 161
QY 303 TTTCCATCAAAAAGAAAGAGGTGTCCAAAACGGTGTTCAGCAAGACTTACGAGAGAGG 362
DB 162 -----GlnArgArgAspArgArgGln----- 168
QY 363 AGGGAGCTGAGCCACCTCAGCCACAGGTACGAGAGGGATGCTGGTGTGCTTGAACCTGG 422
DB 168 yGlyAspAlaGlyGlyGlyGluGlyArg-----LeuGlyProAlaThrProGln 184
QY 423 GCAGCCAGAGATGACATGACGAATCTTCCAGCCAGCGGTCTCCCA----- 471
DB 184 ySerArgArgGlnGlyLeuAlaAlaValThrTptThrGlyGlyGlyProArgProCysGln 204
QY 472 -----ACGCGGAGGAGAAATGTGCCAACTGGTCTGTCTGTA 506
DB 204 yProAlaAlaGlySerValArgSerValArgAspArgSerAlaSerLeuArgThrAl 224
QY 507 GCTAACCAAGGCTGGAGAGTGTGGAGCAGGAGGAGGCCACATGAGGAGGTGACAGCGT 566
DB 224 aThrArgProGlyTptArg-----Gly-ArgArgAlaProSerGlyValArgSerP 242

```


Pred. No.: 0.132 Length: 645
Score: 125.50 Matches: 100
Percent Similarity: 30.99% Conservative: 32
Best Local Similarity: 23.47% Mismatches: 104
Query Match: 6.10% Indels: 190
DB: 12 Gaps: 20

US-10-644-659A-1 (1-1146) x US-10-425-114-45380 (1-645)

QY 9 GGGCGAAAGAAAGCGGGGAGGCGCCAGCCNAGAG---CGCCCTCCGAGATACGCAC 65
DB 93 GYProArgHisGlnGlnGlyProLeuProProLeuGlnProGlyGlyGluAlaHis 112
QY 66 AGCCACCCCTGTCATCAGCTTGGCCGAGGTGGCGAGCTGGCGGAATGAGAACAGCAT 125
DB 113 GlyHisHisLeuHisArgGlnArgLeuAla---GlnGlyGln-----G 128
QY 126 CAGGAGGCCCGAGGAGC-----CTACAGCTGGCTCCGGGAG----- 163
DB 128 lnglyArgProArgAlaAspHisArgAlaValProAspGlnGlnGlyCysArgGluGluA 148
QY 163 ----- 163
DB 148 spProArgProHisArgArgLeuArgArgProArgProLeuProGlyArgValValP 168
QY 164 -----GGACCCAGG-----ACTCACCTCAAGC 185
DB 168 roAlaglyProArgGluGlnGlnGlyGluArgArgAlaValAlaValHis--ArgSe 187
QY 186 TCCTAAACCAATCACACCCCTACTTTCACACAGAAAGCTCAGAGTGCCTCAAGTGC 245
DB 187 rProProAlaValArgProAlaAlaArgGlnArgGlyAspHisProProArgAlaPr 207
QY 246 ACCC-----CGCTGCCAGAGGACAT-----GGAGATGG 275
DB 207 oProGlyArgGluArgGluAspAspHisArgArgProAlaGlyHisArgGlnGlyAspGl 227
QY 276 ACAAGCTCAGAAAGCCCTCAGCTTCTCATCAACAAAGAGAGGTGTCCAAAC 335
DB 227 y-----AlaAlaPro---GlyHisGly---GlnGlnHisValProValHisHis 241
QY 336 -----GGTGTCTCAGCAAGACTTACAGAGAGGAGGAGGAGCTGAGCCACTCAGCCACAG 389
DB 242 ProAlaglyArgGlnGlnAspArgAspGlyArg----- 253
QY 390 GTACGAGAGGATGCTGTGTCTGTTGAACTCGGCGCAGCCAGAAATGACATTGACAGAT 449
DB 253 ----- 253
QY 450 CCTCCACAGCCAGGCTCCCAACGCGGAGGAGAAATGTGCCAACTGTGTCTGAGCT 509
DB 254 ---ProGluHisArgArgAlaAspArgGluGlyArgValArgGlyGlyValProGly 272
QY 510 AACCAAGGCTGGAGAGTGTGAGCGAGGAGGA-----CCACATG----- 551
DB 273 AlaGlnValArgAspArgGluAlaAlaAlaGlyProGluAlaHisLeuArgHisAspArg 292
QY 552 -----GAGGAGTGACAGCTGATAGACACAGAGGAGCGG 584
DB 293 GlyArgArgGluArgArgAlaGlyAlaGluGluGlyGlyHisArgHisArgGlyGlyArg 312
QY 585 CTATGAGGAGGAGGCTGAGGAGGAGCGCGAGGATGGAGTGCAGGTGCTGTGTGTCAG 644
DB 313 -----ArgAspGlyArgGlyProGluArgValGlyHisArgAlaAspGlyAlaArgAla 330
QY 645 GATCAAGGCCCTTGCCTCCAGGTAAACAGATTACAGAGAACTCAACTGCAAGC 704
DB 331 GluArgAsp-----ArgGluArgArgAlaHisGln 340
QY 705 CCAACAGAAATATAGCCAGTGGGCAACTTGAAGGGAGATGGCAGCAGTGGGTGTATGA 764
DB 341 ProArgHisLeuProAlaHisGluGluLeuHisHis----- 352

QY 765 ACACATACAAATCCAGAAAGCTCAATCCTTTCAGTGAAGAGTTTGTATTACGAGTGCCTAT 824
DB 353 -----LeuArgArgValHis 357
QY 825 GTCCACCCGCT-----ACACAAAGGAGATGAGGCTA----- 857
DB 358 HisHisProHisArgAlaGlyLeuProAlaArgArgAlaGlyLeuGluValArgLeuArg 377
QY 858 -----TGGCGCCGCCAAAGAAAGAACCAAAACTGTCTGAAAGGCC 896
DB 378 AlaLeuHisGlyAlaHisHisCysHisProGlnArgHisHisHis----- 393
QY 897 CACGCTGTGAGGAGCAGCATCTACAGGGA----- 926
DB 394 -----AspHisLeuGlnGlyProArgGluAlaValAlaAspAlaArgLeu 408
QY 927 -----AATGATGGACATGTCTTCATTATCTGCACATGGCTCG 965
DB 409 ValGluAlaGlnGlyAspLeuArgHisGlyHisArgAlaArgAspLeuHisGly-AlaAr 428
QY 966 CCAC 969
DB 428 gHis 429

RESULT 12
US-10-425-114-45434
; Sequence 45434, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Tongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 45434
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700237247_FU1.pep
US-10-425-114-45434

Alignment Scores:
Pred. No.: 0.137 Length: 339
Score: 124.50 Matches: 86
Percent Similarity: 36.56% Conservative: 31
Best Local Similarity: 26.88% Mismatches: 106
Query Match: 6.05% Indels: 97
DB: 12 Gaps: 16

US-10-644-659A-1 (1-1146) x US-10-425-114-45434 (1-339)

QY 15 AAAGCAAGCGGGGAGGCGCCAGCAAGAGCGCCCTCCGGAAGATACGCACAGCCACCT 74
DB 63 ArgGlyArgProGlnGluProAlaGlnGluArgArgProValProVal-----HisPro 80
QY 75 GGTCTATCAGCTTGGCCCGAGGTGTGCAGAGTGGCGGAATGAGACACATCAGCGAGC 134
DB 81 GlyHisGluArgValPro-----AlaGlyAlaGlyGlu-----HisGlyGlyAla 95
QY 135 CCAGGAGCTCAGAGCTGTGCTCCGCGGAGGAGCCAGGAGTCACTCAAGTCTCTAAACC 194
DB 96 AlaArgAspHisArgGlnAla-Pro-MetGluProAspAspHisLeuArg----- 111
QY 195 AATCACACCCCTTACTTTCACACAGAAAGCTCAGAGTGTGCCCAAGTGCCTACCCGCT 254

112 -----ArgArgGlyProValProGlnHisAlaHisGlnAlaHis 124
255 GCAGAAGACATGAGATGGACAAAGCTCAGAGAAAGCCCTCAGAGTTCTCACATCAA 314
124 isArgArg-----GlnGluAspHisGlyAspArgLeuHisArgLeuLeuAla-Pro 141
315 AAAGAAAGAGGTGTCAAAA CGTGGTTCAGCAAGACTTA-----CGA 356
142 ArgHisArgAlaArgThrProGlyArgHisAspLeuGlyHisGlyHisGlnProArg 161
357 GAGAGGAGG-----CGAGCTGAGCCCATCTCAGCCACAGGTA 392
162 GluLeuArgAlaArgProSerLeuHisGlnGlnGlyArgArgGlyProGlnAspArgLeu 181
393 CGAGAGGAGTGTGTGTGTCTGAACCTGGCCACCCAGAGATGACATTGACAGAACTCT 452
182 ProArgGly-----ProArgAla 187
453 CCACAGCCACCGCTCCCAACCGGAGAGAGAA----- 484
188 ProArgPro-GlyArgPro-ArgGlyGlyGlnGlyAlaAaArgValValArgLeuArgL 207
485 -----AATGTGCCAACTGGTGTCTGAGCTTAACCAAGGCTGGAG 524
207 euArgGlyArgArgGlnGlyGlnLeuProGlnLeuProArgAlaAaProGluAlaGlyG 227
525 AGTGATGGAGCAGGAGGAGCCACATGGAGAGTGACGCGTAGACACAGACAGACAGCGG 584
227 lu-----AlaArgArgProHis-----ArgLeuArgGlnHisA 238
585 CTATGAGAGAGAGGTGAGGAGAGCCCGACAGAGATGAGTGCGTGTGTCAG 644
238 laVal---GluArgLeuArgArgAlaProArgArgAla-----HisA 252
645 GATCAAGCGCCCTTGCCCTCCC-----AGTAAACAGATTTACAGAGAACTCAACTG 698
252 laGlnValHisProLeuLeuProArgLeuArgLeuArgProGlnGlnArgAlaArgArgA 272
699 CAAAGCCCAACAGAAATATAGCCCTAGTGGGCAACTGAAAGGGAGATGGCAGCAGTGGGC 758
272 xArgProArgArgAspLeu-ProAlaProArgArgArgArgA- 286
759 TGATGAACACATACATCCCAAGCTCAATCTTTCAGTCAAGAGTTTGATTACGAGCT 818
287 -----HisAlaLeuProProArgGlnValLysLysArgArg-----ArgArgLys 301
819 GGCATGTCCACCGGCTACACAAAGGAGATGAGGGCTATGCGCGCC 866
302 LVaAsnIleProCvsValProAlaAlaProAlaValTtpProPro 317

US-10-425-114-70922

Alignment Scores:
 Pred. No.: 0.174 Length: 422
 Score: 123.50 Matches: 85
 Percent Similarity: 33.52% Conserved: 33
 Best Local Similarity: 24.15% Mismatches: 98
 Query Match: 6.00% Indels: 138
 DB: 12 Gaps: 17

US-10-644-659A-1 (1-1146) x US-10-425-114-70922 (1-422)

QY	5	CTCCGGCGCAAAAGAAAGCGGGGAGG-----	31
DB	20	ValArgAlaArgAlaArgGlyArgArgArgCysLeuArgGlyGluLeuVal	39
QY	32	GCCAGCCAGAGGCGCTCCGGGAAGATACGACAGCCACCTCGTGCATCAGCTGGCCCC	91
DB	40	GlnGlnProArgValProAspValArgArgProAlaProPro-----	53
QY	92	GAGGTTGGCAGCAGTGGCGGAATCAGAACAGCATC-----AGGCAGCGCCACG	138
DB	54	-----ArgArgValAlaAlaLeuValLeuArgArgArgGlyVal	66
QY	139	GAGCCTACAGCTGCTCCGGGAGGACCCAGGACTCACCTCAAGCTCTTAACCAATC	198
DB	67	ValProValGlyValProGlyGly-----	75
QY	199	ACACCCCTACTTCACACCAGAAAGCTCAGAGTCCCAAGTCGCCACCCCGCTGCCA	258
DB	76	---ProAlaValArgArgValGlyLeuGlyProGlu-----	92
QY	259	GAAGGACATGAGATGGACAAAGCTCAGAGAAAGCCCTGAGTGTTCACATCAAAAG	318
DB	93	Glu---HisGlyArgHisAlaGlyAlaGlyArgProGlnArgArgArgArg	111
QY	319	AAAGAGGTGTCAAACCGTGTGCAGCAAGCTTACGAGAGAGGGGACGTGAGCCAC	378
DB	112	ArgArgVal-----ArgGlyAlaAspLeuGlnHis	121
QY	379	CTCAGCCACAGGTACGAGAGGATGCTGGTGTG-----CTTGAACCTGGGCGCCAG	432
DB	122	LeuArgHisLeuAlaArgLeuProGlyProProArgLeuProAspValArgProArg	141
QY	433	AATGACATGTGACAG-----ATCTCCACAGCAGCGGTCTCCCAACG	474
DB	142	ArgArgAlaArgArgArgProArgProAspLeuArgHis-AlaGlnAlaProGluArg	161
QY	475	C---GGAGGAGAAATGCGCAACTGTGTCTGAGCTAACCAAGGCTGAGAGTGTATG	531
DB	161	GHisGly-ArgArgArgGlyGlnProGlyVal-----	171
QY	532	GAGCAGGAGGAGCC-----	545
DB	172	--AlaGlyGlyAlaLeuLeuArgArgAspProAspThrArgAsnGluArgArgGlnHisG	191
QY	546	-----CACATGAGAGGTACACGCTGACACAGAGGACAGCGGTATGGAG--	592
DB	191	IuArgGlnGlyHisAlaGluProProHisArgArgHisArgArgGlnArgTrpGlyG	211
QY	593	-----GAGAGGCTGAGGAGGAGCGCCGAGCAGGATGGAGTGC-----	628
DB	211	IntYrProGlnGluGluIuysTyAlaArgAspProGluArgArgGluLeuGlyProProA	231
QY	629	-----AGTGGCTGTGTTCAGGATCAGCGCCCTTGCCTCCAGGTAAACAG	677
DB	231	IaAlaHisGlyArgHisLeuArgSer-----ValProGlnAspValG	245
QY	678	ATTTCACAGAAACTCAACTGCGAAAGCCCAACAGAAATATAGCCCCAGTGGGCAACTGAA	737
DB	245	InValArgProGlyLeuValLeuProProArgGlyVal-ProAlaGlyArgValArg	264
QY	738	AGGGAGATGGCAGCAGTGGGTGTGATGAACATCAATCCAGAAAGCTCACTCTTTTCA	797

Db 265 ArgArgGly----- 267
 QY 798 TGAAGAGTTGATTACGAGCTGGCCATGTCCACCGCCTACACAAAGAGATGAGGG--- 854
 Db 268 -----ValArgLeuGlyHisArgHisProArgGluProValGlnGlyAsp 283
 QY 855 -----CTATGGCCGCCCAAGAGGAC 878
 Db 284 HisLeuLeuAlaProProGluHisArgAsp 293

RESULT 14

US-10-425-114-71620
 ; Sequence 71620, Application US/10425114
 ; Publication No. US20040034988A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Jingdong
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Screen, Steven E
 ; APPLICANT: Tabaska, Jack E
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(5313)B
 ; CURRENT APPLICATION NUMBER: US/10/425,114
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 73128
 ; SEQ ID NO 71620
 ; LENGTH: 460
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: UC-2MROB73024G01_FLI.pep
 US-10-425-114-71620

Alignment Scores:
 Pred. No.: 0.178 Length: 460
 Score: 123.50 Matches: 85
 Percent Similarity: 33.52% Conservative: 33
 Best Local Similarity: 24.15% Mismatches: 98
 Query Match: 6.00% Indels: 138
 DB: 12 Gaps: 17

US-10-644-659A-1 (1-1146) x US-10-425-114-71620 (1-460)

QY 5 CTCGGCGGAAAGGAAAGCGGGAGG----- 31
 Db 58 ValArgAlaArgAlaArgGlyGlyArgArgCysLeuArgGlyGlyGluLeuVal 77
 QY 32 GCCAGCCAAAGCGCCCTCGGAAGATGACGACGCCACCCCTGGTTCATCAGTTGGCCC 91
 Db 78 GlnGlnProArgValProAspValArgArgProAlaProPro----- 91
 QY 92 GAGGTTGGCAGCAGTGGCGAATGAGAACAGCATC-----AGGCGAGGCCAG 138
 Db 92 -----ArgArgValAlaAlaLeuValLeuArgArgArgGlyVal 104
 QY 139 GAGCCTACAGCTGGCTGGGAGGAGGAGCCAGACTCACTCAAGCTCTCAACCAATC 198
 Db 105 ValProValGlyGlyValProGlyGly----- 113
 QY 199 ACACCCCTACTTCACACCAAGAGCTCAGATGCCCAAGTCGCCACCCCGCTGCCA 258
 Db 114 ---ProAlaValArgArgValGlyArgLeuGlyProGlu-----ProGlnGlyPro 130
 QY 259 GAAGCAGTCAGATGACAAAGCTCAGAGAAAGCCCTCAGGTTCTCATCAATAAAAG 318
 Db 131 Gln---HisGlyArgHisAlaGlyArgProGlnArgArgArgArgArg 149
 QY 319 AAAGAGTGTCAAAACGGTGGTCAGCAAGACTTACGAGGAGGAGGCGTGCAGCCAC 378
 Db 150 ArgArgVal-----ArgGlyAlaAspLeuGlnHis 159

QY 379 CTCAGCCACAGGTACGAGGAGGTGCTGTGTG-----CTTGAACCTGGGAGCCAGAG 432
 Db 160 LeuArgHisLeuAlaArgLeuProGlyGlyProProArgLeuProAspValArgProArg 179
 QY 433 AATGACATTGACAGA-----ATCTCCACAGCAGCCGCTCCCAAG 474
 Db 180 ArgArgAlaArgArgArgProArgProAspLeuArgHis-AlaGlnAlaProGluAr 199
 QY 475 C---GGAGGAGAAATGTGCCAACCTGGTCTCTGACGTAAACCAAGGCTGGAGAGTATG 531
 Db 199 GHisGly-ArgArgArgGlyGlnProGlyVal----- 209
 QY 532 GAGCAGGAGGAGCC----- 545
 Db 210 --AlaGlyGlyAlaLeuLeuArgArgAspProAspThrArgAsnGluArgArgGlnHisG 229
 QY 546 -----CACATGGAGAGTGACGGTGTAGACACAGCAGCAGCGCTATGGAG-- 592
 Db 229 luArgGlnGlyHisAlaGluProProHisArgArgHisArgArgGlnArgTrpGlyG 249
 QY 593 -----GAGAGGCTGAGGAGGCGCCGAGCAGGATGAGTGC----- 628
 Db 249 lntyrProGlnGluGluLysTyrAlaArgAspProGluArgArgGluLeuGlyProProA 269
 QY 629 -----AGTGGCTGTGTCAGGATCAAGCGCCCTTGGCTCCAGGTAACAG 677
 Db 269 laAlaHisGlyArgHisLeuArgSer-----ValProGlnAspValG 283
 QY 678 ATTTACAGAGAACTCAACTGCAAGCCCAACAGAAATATAGCCAGTGGCAGCTTGA 737
 Db 283 lntValArgArgProGlyLeuValLeuProProArgGlyVal-ProAlaGlyArgValArg 302
 QY 738 AGGAGATGCGCAGCAGTGGGCTGATGAACACATCAATCCACAGAGCTCAATCCTTTCAG 797
 Db 303 ArgArgGly----- 305
 QY 798 TGAAGAGTTGATTACGAGTGGCCATGTCCACCCGCTACACAAAGAGATGAGGG--- 854
 Db 306 -----ValArgLeuGlyHisArgHisProProArgGluProValGlnGlyAsp 321
 QY 855 -----CTATGGCGCCGCCCAAGAGGAAAC 878
 Db 322 HisLeuLeuAlaProProGluHisArgAsp 331

RESULT 15

US-10-029-386-33391
 ; Sequence 33391, Application US/10029386
 ; Publication No. US20030194704A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FC
 ; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
 ; FILE REFERENCE: AEOMICA-X-2
 ; CURRENT APPLICATION NUMBER: US/10/029,386
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 34288
 ; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 33391
 ; LENGTH: 293
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AC002104.1
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
 ; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 1.4
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
 US-10-029-386-33391

Alignment Scores:

Pred. No.: 0.213 Length: 293
Score: 122.00 Matches: 70
Percent Similarity: 28.15% Conservative: 15
Best Local Similarity: 23.18% Mismatches: 99
Query Match: 5.93% Indels: 118
DB: 14 Gaps: 12

Search completed: May 4, 2004, 08:59:57
Job time : 129.5 secs

US-10-644-659A-1 (1-1146) x US-10-029-386-33391 (1-293)

```
QY 62 GCACAGCCACCTGGTCATCAGCT----- 85
Db 7 AlalysTrpProTyrGlySerArgArgLeuAspProGlyAlaArgThrGlyAlaLys 26
QY 86 TGGCCCGAGGTGGCAGCAGTGGCGAATGAGAACAGCATCAGCCAGG----- 133
Db 27 TrpProGlySerGlySerGlyLysArgArgProAlaArgGlyArgCysProGlyPro 46
QY 134 -----CCCAGGAGCCTACAGGCT 151
Db 47 ThrGlnLeuProProAlaProArgAlaProAlaProArgProSerArgAlaLeuGlnPro 66
QY 152 GGCTGCGGGAGGAGCCAGGAGCTCACCTCAAGCTCCTAAACCAATCACACCCCTACTT 211
Db 67 -----ArgProGlyPro-----AlaGlnArgProArgProLeuPhe 78
QY 212 CACACGAGAAAGCTCAGA-----GTGCCCCAAAGTCGCCACCCC 250
Db 79 HisArgArgProAlaArgArgGluAlaGlnGlyCysAspProArgArgAlaHisPro 98
QY 251 GCCTGCCAGAGGACATGGATG-----GACAAAGCTCAGAGAAAGCCC 295
Db 99 AlaArgGlnAlaProTyrGluMetArgGlyHisSerGlnGlyProAlaGluArgAlaGln 118
QY 296 CTGAGGTTCTCACAATCAAAAGAGGTGTCAAAACGGTGGTCAGCAGACTTACG 355
Db 119 ArgArgSerLeuProHisProArgProArgProArgAsnArgArgThrArgAlaVal 138
QY 356 AGAGAGGAGGAGCGTGCAGCCACCTCAGCCACAGCATCAGAGGAGGATGCTGGTGGCTTG 415
Db 139 ProAlaGlyGlyGlyAlaThrGlnAlaGluGlyThrLeuAlaAsnPheAlaProPro 158
QY 416 AACCTGGGCGAGCCAG----- 430
Db 159 GluValAlaSerArgArgProGlyAlaHisProProAlaAlaArgProArgProGlyAla 178
QY 431 -----AGAATGACATTGACA 445
Db 179 ArgProProGlnProGlyArgLeuGlyArgArgCysAspLeuArgValGlnSerAla 198
QY 446 GAATCCTCCACAGCCACGCTCCCAACGCGA-----GGAGAAAATGTCCCAACCTGG 499
Db 199 ArgArgAlaAlaArgThrProProArgGlyProProAlaArgArgAlaLeuThrTrp 218
QY 500 -----TGT 502
Db 219 AspIleCysGlyArgLysLeuMetSerLeuArgSerMetGluProGlyGluArgLeuCys 238
QY 503 CTGAGCTTACCAAGGCTGGAGAGTGATGGAGCAGGAGGCCACATGGAGAGTGACA 562
Db 239 SerSer-----TrpHisArgArgThrProSerArgSerAlaCys 251
QY 563 GCGTAGACACAGAGGACCGGCTATGGAG-----GAGAGGCTG 601
Db 252 AlaArgSerLysThrThrAlaGluGlyHisThrThrArgTrpLeuGlyGlyArgThrLeu 271
QY 602 AGGAGAGGCCCGAGCAGGATGGAGTGCAGGTGGCTGGTCAGGATCAAGCGCCCTTGC 661
Db 272 GlnSerMetSerGlnArgProHisCysArgHisCys-----SerMetAlaProAla 288
QY 662 CCTCCC 667
Db 289 GlyPro 290
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 4, 2004, 08:49:07 ; Search time 64 Seconds
(without alignments)
10118.733 Million cell updates/sec

Title: US-10-644-659A-1
Perfect score: 2058
Sequence: 1 aggttcggcggaagaagga.....tgattacgctactcaagtga 1146

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p model -DSV=xlp
-Q/cgn2_1/USFTO.spool_p/US10644659/runat_04052004_084901_28500/app_query.fasta_1.1287
-DB=A_Geneseq_29Jan04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USR=US10644659 @CGN 1.1 81 @runat_04052004_084901_28500 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_29Jan04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2029	98.6	381	7	ADB65015 Human pro
2	1988	96.6	456	4	ABG27995 Novel hum
3	647	31.4	121	4	AAM21624 Peptide #
4	647	31.4	121	4	ABB43984 Peptide #
5	647	31.4	121	4	AAM37926 Peptide #
6	647	31.4	121	4	ABB26893 Protein #
7	647	31.4	121	4	AAM77709 Human bon
8	647	31.4	121	4	AAM64987 Human bra
9	647	31.4	121	4	ABG53364 Human liv
10	362	17.6	69	3	AAG00337 Human sec

11	362	17.6	70	2	RAY11623
12	258	12.5	399	4	ABB60049
13	205	10.0	183	4	ABB58876
14	190	9.2	37	4	AAM21625
15	190	9.2	37	4	ABB43985
16	190	9.2	37	4	AAM37927
17	190	9.2	37	4	ABB26894
18	190	9.2	37	4	AAM77710
19	190	9.2	37	4	AAM64988
20	190	9.2	37	4	ABG53365
21	136	6.6	1938	6	ABB9398
22	134	6.5	1938	6	ABP76682
23	133.5	6.5	83	3	ABAB32055
24	133.5	6.5	161	3	ABG58353
25	132	6.4	956	3	ABO53053
26	131.5	6.4	95	3	ABG58354
27	130.5	6.3	1739	6	ABG36105
28	130.5	6.3	1739	6	ABG36106
29	128.5	6.2	1017	4	ABAB59813
30	128.5	6.2	1615	4	ABAB59826
31	128	6.2	1938	6	ABP76680
32	127	6.2	1938	6	ABP76679
33	125.5	6.1	563	4	ABG30256
34	124.5	6.0	87	3	ABG58854
35	123	6.0	1267	4	ABAB61050
36	121.5	5.9	440	4	ABG05146
37	121	5.9	926	4	ABAB65135
38	121	5.9	1182	5	AAE19798
39	121	5.9	1182	6	ABU62544
40	120.5	5.9	1019	4	ABAB1928
41	120.5	5.9	1714	4	ABG08417
42	120.5	5.9	1714	4	ABG02138
43	120.5	5.9	2803	5	ABAB08161
44	120	5.8	485	5	ABP79905
45	120	5.8	900	4	ABG03533

ALIGNMENTS

RESULT 1

ADB65015

ID ADB65015 standard; protein; 381 AA.

XX ADB65015;

AC ADB65015;

DT 04-DEC-2003 (first entry)

DE Human protein encoded by clone SKMUS20026340.

XX Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;
cell regeneration; membrane protein; signal transduction-related protein;
transcription-related protein; osteoporosis; neurological disease;
cancer; tumour.

XX Homo sapiens.

OS Homo sapiens.

PN EP1308459-A2.

XX EP1308459-A2.

PD 07-MAY-2003.

XX 28-MAR-2002; 2002EP-00007401.

XX 05-NOV-2001; 2001JP-00379298.

XX 25-JAN-2002; 2002JP-00350978.

PR (HELI-) HELIX RES INST.

XX (HELI-) HELIX RES INST.

PA (HELI-) RES ASSOC BIOTECHNOLOGY.

XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;

PI Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuho Y;

XX WPI; 2003-450961/43.

DR WPI; 2003-450961/43.

DR N-PSDB; ADB63045.
 XX New polynucleotides and polypeptides, useful for developing a diagnostic
 PT marker or medicines for regulation of their expression and activity, or
 PT as targets of gene therapy.
 XX
 PS Claim 1; Page; 222pp; English.
 XX

The invention discloses a polynucleotide comprising a sequence selected
 CC from 1970 fully defined nucleotide sequences which encode novel
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
 CC or its partial peptide, an antibody binding to the polypeptide or peptide
 CC of the polynucleotide, immunologically assaying the polypeptide or
 CC peptide of the polynucleotide by contacting the polypeptide or peptide
 CC with the antibody of the encoded protein, and observing the binding
 CC between the two, a transformant carrying the polynucleotide in an
 CC expressible manner and an antisense polynucleotide. The oligonucleotide
 CC is useful as a primer for synthesizing the polynucleotide, or as a probe
 CC for detecting the polynucleotide. The polynucleotides and encoded
 CC proteins are useful as pharmaceutical agents and many disease-related
 CC genes may be included in them, for developing a diagnostic marker or
 CC medicines for regulation of their expression and activity, or as targets
 CC of gene therapy. The genes are involved in tissue and/or cell
 CC regeneration. Membrane proteins, signal transduction-related proteins,
 CC transcription-related proteins, disease-related proteins and genes
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
 CC neurological diseases, cancer, tumors). The cDNA may be used to regulate
 CC the activity or expression of the encoded protein to treat diseases. The
 CC sequence presented is a protein of the invention. Note: Some of the
 CC sequence data for this patent is not represented in the printed
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office.
 XX
 SQ Sequence 381 AA;

Alignment Scores:
 Pred. No.: 5, 25e-181 Length: 381
 Score: 2029.00 Matches: 381
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 98.59% Indels: 0
 DB: 7 Gaps: 0

US-10-644-659A-1 (1-1146) X ADB65015 (1-381)

QY	1	ATGGCTCCGGCGGAAAGAAAGCGGGAGGCGCCAGCCAGAGCGCCCTCCGGAAGATA	60
DB	1	MetAlaProGlyGluLysSerGlyGluGlyProAlaLysSerAlaLeuArgLysIle	20
QY	61	CGCAGACCCACCTGGTTCATCAGCTTGGCCGAGGTTGGCAGCTGGCGGAATGAGAAC	120
DB	21	ArgThrAlaThrLeuValIleSerLeuAlaArgGlyTrpGlnGlnTrpAlaAsnGluAsn	40
QY	121	AGCATCAGCAGCGCCAGAGCGCTACAGCTGGCTGCCGGAGGAGCCAGGACTCACT	180
DB	41	SerIleArgGlnAlaGlnGluProThrGlyTrpLeuProGlyGlyThrGlnAspSerPro	60
QY	181	CAGGCTCTTAAACCATCACCCCTACTTCCACACAGAAAGCTCAGAGTGGCCCAAG	240
DB	61	GlnAlaProGlyProIleThrProThrSerHisGlnLysAlaGlnSerAlaProLys	80
QY	241	TCGCCACCCCGCTCCAGAGGACATGAGATGGACAAAGCTCAGAGAAAGCCCTGAG	300
DB	81	SerProProArgLeuProGluGlyHisGlyAspGlyGlnSerSerGluLysAlaProGlu	100
QY	301	GTTCCTCACATCAAAAGAAAGAGGTGTCCAAACGGTGGTCAGCAAGACTTACAGAGA	360
DB	101	ValSerHisIleLysLysLysGluValSerLysThrValValSerLysThrTrpGluArg	120
QY	361	GGAGGGGAGCTGAGCCACCTCAGCCACAGGTACGAGAGGAGTGGTGGTGGTGAACCT	420
DB	121	GlyGlyAspValSerHisLeuSerHisArgTrpGluArgAspAlaGlyValLeuGluPro	140

QY	421	GGCAGCCAGAGAAATGACATTGACAGAACTCTCCACAGCCAGCGCTCCCAAGCGGAGG	480
DB	141	GlyGlnProGluAsnAspIleAspArgIleLeuHisSerHisGlySerProThrArgArg	160
QY	481	AGAAAATGTGCCAACCTGTGTCTGAGCTTAACCAAGGCTGGAGAGTGTATGAGCAGAG	540
DB	161	ArgLysCysAlaAsnLeuValSerGluLeuThrLysGlyTrpArgValMetGluGlnGlu	180
QY	541	GAGCCCATGAGGAGGTGACAGCCTAGACAGAGGAGCGGCTATGAGGAGAGGCT	600
DB	181	GluProThrTrpArgSerAspSerValAspThrGluAspSerGlyTrpGlyGlyGluAla	200
QY	601	GAGGAGAGCCCGCAGCAGGATGAGTGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	660
DB	201	GluGluArgProGluGlnAspGlyValGlnValAlaValValArgLysArgProLeu	220
QY	661	CCCTCCAGGTAAACAGATTACAGAGAACTCACTGCAAGCCCAAGCCCAAGAAATATAGC	720
DB	221	ProSerGlnValAsnArgPheThrGluLysLeuAsnCysLysAlaGlnGlnLysTrpSer	240
QY	721	CCAGTGGGCAACTTGAAGGGAGATGGCAGCAGTGGGCTGATGAACACACATCAATCCCA	780
DB	241	ProValGlyAsnLeuLysGlyArgTrpGlnGlnTrpAlaAspGluHisGlnSerGln	260
QY	781	AGCTCAATCTTTTACGTGAAGATTGATTACAGAGTGGCCATGTCCACCGCTACAC	840
DB	261	LysLeuAsnProPheSerGluGluPheAspTrpGluLeuAlaMetSerThrArgLeuHis	280
QY	841	AAAGGAGATGAGGGCTTATGGCCGCCCCCAAGAGGAACCAAACTGCTGAAAGGCCAAG	900
DB	281	LysGlyAspGluGlyTrpGlyArgProLysGluGlyThrLysThrAlaGluArgAlaLys	300
QY	901	CGTCTGAGGAGACATCTACAGGGAATGATGGACATGTGCTTCTATTCATTCGCAATG	960
DB	301	ArgAlaGluGluHisIleTrpArgGluMetMetAspMetCysPheIleCysThrMet	320
QY	961	GCTCGCCACAGCAGAGATGCGAAGATCCAGGTACTTTTGGAGATCTCTTTACAGATAC	1020
DB	321	AlaArgHisArgArgAspGlyLysIleGlnValThrPheGlyAspLeuPheAspArgTrp	340
QY	1021	GTTCGTATTTCAGATAAAGTACGTGGCATTCATCGCTGCGCAGGAACATCGACTGGTA	1080
DB	341	ValArgIleSerAspLysValValGlyIleLeuMetArgAlaArgLysHisGlyLeuVal	360
QY	1081	GACTTTGAGGAGAGATGCTATGCGAAGCCGAGATGACCATGTTGTGATTACGCTACTC	1140
DB	361	AspPheGluGlyGluMetLeuTrpGlnGlyArgAspAspHisValIleThrLeuLeu	380
QY	1141	AAG 1143	
DB	381	Lys 381	

RESULT 2
 ABG27995
 ID ABG27995 standard; protein; 456 AA.
 XX
 AC ABG27995;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #27996.
 XX
 DE Human, chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 XX W0200175067-A2.
 PN
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX

PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 PA (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 DR N-PSDB; AAS92182.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 58354; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (II) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG3077 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 456 AA;

Alignment Scores:
 Pred. No.: 3.9e-177 Length: 456
 Score: 1988.00 Matches: 381
 Percent Similarity: 86.00% Conservative: 0
 Best Local Similarity: 86.00% Mismatches: 0
 Query Match: 96.60% Indels: 62
 DB: 4 Gaps: 1

US-10-644-659A-1 (1-1146) x ABG27995 (1-456)

QY 1 ATGGCTCCGGCGAAAGGAGCGGGAGGCGCCAGAGCGCCCTCCGGAAGATA 60
 DB 14 MetAlaProGlyGluLysGluSerGlyGluGlyProAlaLysSerAlaLeuArgLysile 33
 QY 61 CGCAGAGCCACCTGGTTCATCAGCTTGGCCCGAGGTTGGCAGAGTGGGGGAAATGAGAC 120
 DB 34 ArgThrAlaThrLeuValIleSerLeuAlaArgGlyTrpGlnGlnTrpAlaAsnGluAsn 53
 QY 121 AGCATCAGGAGGCGCCAGGAGCTACAGGCTGGCTGCCGGAGGAGCCAGGACTCACCT 180
 DB 54 SerIleArgGlnAlaGlnGluProThrGlyTrpLeuProGlyGlyThrGlnAspSerPro 73
 QY 181 CAAGCTCCTAAACCAATCAGACCCCTACTTCACACAGAAAGCTCAGAGTGGCCCAAG 240
 DB 74 GlnAlaProLysProIleThrProProThrSerHisGlnLysAlaGlnSerAlaProLys 93
 QY 241 TCGCCACCCCGCTCCAGAGAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCTGAG 300
 DB 94 SerProProArgLeuProGluGlyHisGlyAspGlyGlnSerSerGluLysAlaProGlu 113
 QY 301 GTTCTCATCAAAAGAGAGGTGTCCAAAGCGGTGTCCAGCAGACTTACGAGAGA 360

RESULT 3
 AM21624
 ID ARM21624 standard; protein; 121 AA.

DB 114 ValSerHisIleLysLysLysGluValSerLysThrValValSerLysThrTyrGluArg 133
 QY 361 GGAGGGCAGCTGAGCCACCTCAGCCACAGGTACGAGAGGAGTGTGGTGTGTTGAACCT 420
 DB 134 GlyGlyAspValSerHisLeuSerHisArgTyrGluArgAspAlaGlyValLeuGluPro 153
 QY 421 GGGCAGCCAGAGATGACATTCACAGATCCCTCCACAGCCAGCGGTCCCAACGCGGAGG 480
 DB 154 GlyGlnProGluAsnAspIleAspArgIleLeuHisSerHisGlySerProThrArgArg 173
 QY 481 AGAAATGTCCCAACCTGTGTGTGAGCTAAACCAAGGCTGGAGAGTGTGAGCAGGAG 540
 DB 174 ArgLysCysAlaAsnLeuValSerGluLeuThrLysGlyTrpArgValMetGluGlnGlu 193
 QY 541 GAGCCACATGGAGGAGTGCAGCGTAGACACAGAGGACAGCGGTATCAGAGGAGGCT 600
 DB 194 GluProThrTrpArgSerAspSerValAspPheGluAspSerGlyTyrGlyGlyGluAla 213
 QY 601 GAGGAGGCGCCAGCAGGATGAGTGCAGGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 660
 DB 214 GluGluArgProGluGlnAspGlyValGlnValAlaValAlaValArgIleLysArgProLeu 233
 QY 661 CCCTCCAG----- 669
 DB 234 ProSerGlnGlyGlyCysLeuHisGluGluGluLysTyrSerLeuTyrArgLysIleTyr 253
 QY 669 ----- 669
 DB 254 GlyValGlySerGluSerProPheSerLeuSerProTyrArgSerGluThrGluGluLys 273
 QY 669 ----- 669
 DB 274 IleMetProLeuGlyValProCysIleLeuLysLeuProLeuGlyAsnGlnIleSerGly 293
 QY 670 -----GTAACAGATTACAGAGAACTCAACTGCAAGAGCCCAACAGAGAA 714
 DB 294 SerGluPheGlyTrpValAsnArgPheThrGluLysLeuAsnCysLysAlaGlnLys 313
 QY 715 TATAGCCAGTGGCGCACTTGAAGGAGAGATGCGCAGTGGGCTGATCAACACATACAA 774
 DB 314 TyrSerProValGlyAsnLeuLysGlyArgTrpGlnGlnTrpAlaAspGluHisIleGln 333
 QY 775 TCCAGAGCTCAATCTTTCAGTGAAGATTGTGATTACGAGTGGCCATGTCCACCCGC 834
 DB 334 SerGlnLysLeuAsnProPheSerGluGluPheAspTyrGluLeuAlaMetSerThrArg 353
 QY 835 CTACACAAAGAGAGTGGGCTATGGCCGCCCAAGAGAGGACCAAACTCTGCTGAAGG 894
 DB 354 LeuHisLysGlyAspGluGlyTyrGlyArgProLysGluGlyThrLysThrAlaGluArg 373
 QY 895 GCCAAGCGTCTGAGGAGCACATCTACAGGAAATGATGACATGTCTCATTAATCTGC 954
 DB 374 AlalysArgAlaGluGluHisIleTyrArgGluMetMetAspMetCysPheIleLysCys 393
 QY 955 ACATGCTCGCCACAGAGAGATGCAAGATCCAGTACTTTTGGAGATCTCTTTGAC 1014
 DB 394 ThrMetAlaArgHisArgArgAspGlyLysIleGlnValThrPheGlyAspLeuPheAsp 413
 QY 1015 AGATACCTTCGTATTTTCAGATAAAGTAGTGGGCTTCTCATGCTGCCAGGAAACATGA 1074
 DB 414 ArgTyrValArgIleSerAspLysValValGlyIleLeuMetArgAlaArgLysHisGly 433
 QY 1075 CTGTGACCTTGAAGGAGAGATGCTATGCGAGGCGGAGATGACCATGTGTGTGTGTGTGT 1134
 DB 434 LeuValAspPheGluGlyGluMetLeuTrpGlnGlyArgAspAspHisValValIleThr 453
 QY 1135 CTACTCAAG 1143
 DB 454 LeuLeuLys 456


```

XX AC AAM21624;
XX DT 12-OCT-2001 (first entry)
XX DE Peptide #8058 encoded by probe for measuring cervical gene expression.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer.
XX OS Homo sapiens.
XX PN WC2000157278-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000670.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX WPI; 2001-488901/53.
XX DR Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human cervical epithelial cells.
XX PS Claim 27; SEQ ID NO 26450; 487pp; English.
XX CC The present invention relates to human single exon nucleic acid probes
XX CC (SENPs: see AAI10068-AA128459). The present sequence is a peptide encoded
XX CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
XX CC can be used to produce a single exon microarray, which can be used for
XX CC measuring human gene expression in a sample derived from human cervical
XX CC epithelial cells. By measuring gene expression, the probes are therefore
XX CC useful in grading and/or staging of diseases of the cervix, notably
XX CC cervical cancer. Note: The sequence data for this patent did not form
XX CC part of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 121 AA;

Alignment Scores:
Pred. No.: 1.04e-51 Length: 121
Score: 647.00 Matches: 121
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.44% Indels: 0
DB: 4 Gaps: 0

US-10-644-659A-1 (1-1146) x AAM21624 (1-121)
QY 262 GGACATGGATGACAAAGCTCAGAGAAAGCCCTGAGTTCTTCACATCAAAAGAA 321
Db 1 GlyHisGlyAspGlyGlnSerSerGluLysAlaProGluValSerHisLeuLysLys 20
QY 322 GAGGTGTCCAAAACGGTGTTCACAAAGACTTACGAGAGAGGGGACGTGAGCCACCTC 381
Db 21 GluValSerLysThrValValSerLysThrThrGluArgGlyGlyAspValSerHisLeu 40
QY 382 AGCCACAGGTACGAGAGGGATGCTGTGTGCTTGAACCTGGCCAGCCAGAGATGACATT 441
Db 41 SerHisArgTyrGluArgAspAlaGlyValLeuGluProGlyGlnProGluAsnAspIle 60
QY 442 GACAGATCTCTCCACAGCCACCGCTCCCAACCGGAGGAGAAATGTGCCAACCTGGTG 501

```

```

Db 61 AspArgIleLeuHisSerHisGlySerProThrArgArgLysCysAlaAsnLeuVal 80
QY 502 TCTGAGCTAACCAAGGCTGAGAGTGTGAGAGCAGGAGCCACATGAGAGGTGAC 561
Db 81 SerGluLeuThrLysGlyTyrArgValMetGluGlnGluProThrTyrPargSerAsp 100
QY 562 AGCGTAGACACAGAGCAGCGGCTATGGAGGAGAGCTGAGGAGAGCCCGAGCAGGAT 621
Db 101 SerValAspThrGluAspSerGlyTyrGlyGlyAlaGluGluAspProGluGlnAsp 120
QY 622 GGA 624
Db 121 Gly 121

RESULT 4
ABB43984
ID ABB43984 standard; peptide; 121 AA.
XX AC ABB43984;
XX DT 04-FEB-2002 (first entry)
XX DE Peptide #11490 encoded by human foetal liver single exon probe.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX OS Homo sapiens.
XX PN WC2000157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000669.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX WPI; 2001-483447/52.
XX DR Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human fetal liver.
XX PS Claim 27; SEQ ID NO 36619; 639pp + Sequence Listing; English.
XX CC The invention relates to a single exon nucleic acid probe for measuring
XX CC human gene expression in a sample derived from human foetal liver. The
XX CC single exon nucleic acid probes may be used for predicting, measuring and
XX CC displaying gene expression in samples derived from human fetal liver. The
XX CC present sequence is a peptide encoded by a single exon nucleic acid probe
XX CC of the invention. Note: The sequence data for this patent did not form
XX CC part of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 121 AA;

Alignment Scores:
Pred. No.: 1.04e-51 Length: 121
Score: 647.00 Matches: 121
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.44% Indels: 0
DB: 4 Gaps: 0

```

US-10-644-659A-1 (1-1146) x ABB43984 (1-121)

QY 262 GGACATGGAGATGGACAAAGCTCAGAGAAAGCCCTGAGGTTTCTCACATCAAAAGAAA 321
 Db 1 GlyHisGlyAspGlyGlnSerSerGluLysAlaProGluValSerHisIleLysLys 20
 QY 322 GAGGTGTCCAAACCGTGTGTCAGAGACTTACGAGAGAGAGGGGAGTGCACCTC 381
 Db 21 GluValSerLysThrValValSerLysThrTyrGluArgGlyGlyAspValSerHisLys 40
 QY 382 AGCCACAGGTACGAGAGGATGCTGTGCTTGAACCTGGCAGCCAGAGATGACATT 441
 Db 41 SerHisArgTyrGluArgAspAlaGlyValLysGluProGlyGlnProGluAsnAsp 60
 QY 442 GACAGAACTCTCCACAGCCAGCGCTCCCAACGGGAGGAGAGAAATGTGCCACCTG 501
 Db 61 AspArgIleLeuHisSerHisGlySerProThrArgArgLysCysAlaAsnLeuVal 80
 QY 502 TCTGAGCTAACCAAGGCTGGAGAGTGTGAGCAGGAGGAGCCACATGGAGGAGTGAC 561
 Db 81 SerGluLeuThrLysGlyTrpArgValMetGluGlnGluProThrTrpArgSerAsp 100
 QY 562 AGCGTAGACACAGAGGAGCGCGCTATGAGGAGAGGCTGAGAGAGGCGCCAGCAGGAT 621
 Db 101 SerValAspThrGluAspSerGlyTyrGlyGluAlaGluArgProGluGlnAsp 120
 QY 622 GGA 624
 Db 121 Gly 121

RESULT 5

AAM37926
 ID AAM37926 standard; protein; 121 AA.

XX AC AAM37926;

XX DT 17-OCT-2001 (first entry)

DE Peptide #11963 encoded by probe for measuring placental gene expression.
 DE Probe; microarray; human; placenta; antenatal diagnosis;
 KW Genetic disorder.

XX OS Homo sapiens.

XX PN WO200157272-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US0000663.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488897/53.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
 XX gene expression in human placenta.

XX PS Claim 27; SEQ ID NO 38195; 654pp; English.

XX CC The present invention relates to single exon nucleic acid probes (SENP:
 XX see AAI31315-AAI57546). The present sequence is a peptide encoded by one
 XX such probe. The probes are useful for producing a microarray for

CC Predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders

XX SQ Sequence 121 AA;

Alignment Scores:
 Pred No.: 1,04e-51 Length: 121
 Score: 647.00 Matches: 121
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 31.44% Indels: 0
 DB: 4 Gaps: 0

US-10-644-659A-1 (1-1146) x AAM37926 (1-121)

QY 262 GGACATGGAGATGGACAAAGCTCAGAGAAAGCCCTGAGGTTTCTCACATCAAAAGAAA 321
 Db 1 GlyHisGlyAspGlyGlnSerSerGluLysAlaProGluValSerHisIleLysLys 20
 QY 322 GAGGTGTCCAAACCGTGTGTCAGAGACTTACGAGAGAGAGGGGAGTGCACCTC 381
 Db 21 GluValSerLysThrValValSerLysThrTyrGluArgGlyGlyAspValSerHisLys 40
 QY 382 AGCCACAGGTACGAGAGGATGCTGTGCTTGAACCTGGCAGCCAGAGATGACATT 441
 Db 41 SerHisArgTyrGluArgAspAlaGlyValLysGluProGlyGlnProGluAsnAsp 60
 QY 442 GACAGAACTCTCCACAGCCAGCGCTCCCAACGGGAGGAGAGAAATGTGCCACCTG 501
 Db 61 AspArgIleLeuHisSerHisGlySerProThrArgArgLysCysAlaAsnLeuVal 80
 QY 502 TCTGAGCTAACCAAGGCTGGAGAGTGTGAGCAGGAGGAGCCACATGGAGGAGTGAC 561
 Db 81 SerGluLeuThrLysGlyTrpArgValMetGluGlnGluProThrTrpArgSerAsp 100
 QY 562 AGCGTAGACACAGAGGAGCGCGCTATGAGGAGAGGCTGAGAGAGGCGCCAGCAGGAT 621
 Db 101 SerValAspThrGluAspSerGlyTyrGlyGluAlaGluArgProGluGlnAsp 120
 QY 622 GGA 624
 Db 121 Gly 121

RESULT 6

ABB26893
 ID ABB26893 standard; protein; 121 AA.

XX AC ABB26893;

XX DT 23-JAN-2002 (first entry)

XX DE Protein #8992 encoded by probe for measuring heart cell gene expression.
 XX Human; gene expression; heart; microarray; vascular system;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease.

XX OS Homo sapiens.

XX PN WO200157274-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US0000666.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-48899/53.
 XX Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts.
 PT
 XX Claim 15; SEQ ID NO 28663; 530pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart (see
 CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
 CC probe. The probes may be used for predicting, measuring and displaying
 CC gene expression in samples derived from the human heart via microarrays.
 CC By measuring gene expression, the probes are useful for predicting,
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the
 CC human heart and vascular system e.g. cardiovascular disease,
 CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 121 AA;

Alignment Scores:
 Pred. No.: 1.04e-51 Length: 121
 Score: 647.00 Matches: 121
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 31.44% Indels: 0
 DB: 4 Gaps: 0

US-10-644-659A-1 (1-1146) x ABB26893 (1-121)

QY 262 GGACATGGAGATGGACAAAGCTCAGAGAAAGCCCTGAGGTTTCTCATCAAAAGAAA 321
 Db 1 GlyHisGlyAspGlyGlnSerSerGluValSerGluValSerHisLeuLysLys 20
 QY 322 GAGGTCTCAAAACGGTGTCTACGAGACTTACGAGAGAGGAGGAGCGTGCACCTC 381
 Db 21 GluValSerLysThrValValSerLysThrThrGluArgGlyGlyAspValSerHisLeu 40
 QY 382 AGCCACAGGTACGAGAGGATGCTGCTGTGCTTGAACCTGGCAGCCAGAGAAATGACATT 441
 Db 41 SerHisArgTyrGluArgAspAlaGlyValLeuGluProGlyGlnProGluAsnAspIle 60
 QY 442 GACAGAACTCTCCACAGCCAGGCTCCCAACGGAGGAGAGAAATGTGCCAATGCTGGTG 501
 Db 61 AspArgIleLeuHisSerHisGlySerProThrArgArgArgGlyCysAlaAsnLeuVal 80
 QY 502 TCTGAGCTAACCAAGGCTGGAGAGTGTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 561
 Db 81 SerGluLeuThrLysGlyTyrArgValMetGluGlnGluProThrTyrArgSerAsp 100
 QY 562 AGCGTAGACACAGAGAGAGCGGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 621
 Db 101 SerValAspThrGluAspSerGlyTyrGlyGlyGluAlaGluGluArgProGluGlnAsp 120
 QY 622 GGA 624
 Db 121 Gly 121

RESULT 7
 AAM77709
 ID AAM77709 standard; protein; 121 AA.
 XX
 AC AAM77709;
 XX
 DT 06-NOV-2001 (first entry)
 XX

DE Human bone marrow expressed probe encoded protein SEQ ID NO: 38015.
 XX Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma.
 XX
 XX Homo sapiens.
 XX
 PN WO200157276-A2.
 XX
 XX 09-AUG-2001.
 PD
 XX
 XX 30-JAN-2001; 2001WO-US000668.
 PF
 XX 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-0052366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-48899/53.
 XX
 XX Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human bone marrow.
 PT
 XX Example 4; SEQ ID NO 38015; 658pp + Sequence Listing; English.
 XX
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a
 CC protein encoded by one of the probes of the invention
 CC
 XX Sequence 121 AA;

Alignment Scores:

Pred. No.: 1.04e-51 Length: 121
 Score: 647.00 Matches: 121
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 31.44% Indels: 0
 DB: 4 Gaps: 0

US-10-644-659A-1 (1-1146) x AAM77709 (1-121)

QY 262 GGACATGGAGATGGACAAAGCTCAGAGAAAGCCCTGAGGTTTCTCATCAAAAGAAA 321
 Db 1 GlyHisGlyAspGlyGlnSerSerGluValSerGluValSerHisLeuLysLys 20
 QY 322 GAGGTCTCAAAACGGTGTCTACGAGACTTACGAGAGAGGAGGAGCGTGCACCTC 381
 Db 21 GluValSerLysThrValValSerLysThrThrGluArgGlyGlyAspValSerHisLeu 40
 QY 382 AGCCACAGGTACGAGAGGATGCTGCTGTGCTTGAACCTGGCAGCCAGAGAAATGACATT 441
 Db 41 SerHisArgTyrGluArgAspAlaGlyValLeuGluProGlyGlnProGluAsnAspIle 60
 QY 442 GACAGAACTCTCCACAGCCAGGCTCCCAACGGAGGAGAGAAATGTGCCAATGCTGGTG 501
 Db 61 AspArgIleLeuHisSerHisGlySerProThrArgArgArgGlyCysAlaAsnLeuVal 80
 QY 502 TCTGAGCTAACCAAGGCTGGAGAGTGTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 561
 Db 81 SerGluLeuThrLysGlyTyrArgValMetGluGlnGluProThrTyrArgSerAsp 100
 QY 562 AGCGTAGACACAGAGAGAGCGGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 621
 DB: 101 SerValAspThrGluAspSerGlyTyrGlyGlyGluAlaGluGluArgProGluGlnAsp 621

Db 101 SerValAspThrGluAspSerGlyTyrGlyGluAlaGluGluA:ProGluGlnAsp 120

QY 622 GGA 624

Db 121 Gly 121

RESULT 8

AA64987

ID AA64987 standard; protein; 121 AA.

AC AA64987;

DT 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe encoded protein SEQ ID NO: 37092.

DE Human; brain expressed exon; gene expression analysis; probe; microarray;

KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.

XX

OS Homo sapiens.

XX

PN WO200157275-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US000667.

XX

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

PD WPI; 2001-483446/52.

DR

XX

PT Single exon nucleic acid probes for analyzing gene expression in human brains.

XX

PS Example 4; SEQ ID NO 37092; 650pp + Sequence Listing; English.

XX

CC The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is a protein encoded by one of the probes of the invention

XX

SQ Sequence 121 AA;

Alignment Scores:

Pred. No.:	1.04e-51	Length:	121
Score:	647.00	Matches:	121
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	31.44%	Indels:	0
DB:	4	Gaps:	0

US-10-644-659A-1 (1-1146) x AA64987 (1-121)

QY 262 GGACATGAGATGACAAAGCTCAGAGAGCCCTGAGGTTCTCATCAAAAGAAA 321

Db 1 GlyHisGlyAspGlyGlnSerSerGluLysIleProGluValSerHisIleLysLysLys 20

QY 322 GAGGTGTCCAAAACGGTGGTCCAGCAAGACTTACAGAGAGAGGGGCGTGGAGCCACTC 381

Db 21 GluValSerLysThrValValSerLysThrTyrGluArgGlyGlyAspValSerHisLeu 40

382 AGCCAGAGTACAGAGAGGATCTGCTGTCTTGAACCTGGCCAGCAGAGATGACATT 441

Db 41 SerHisArgTyrGluArgAspAlaGlyValLeuGluProGlyGlnProGluAsnAspIle 60

QY 442 GACAGAAATCTCCACAGCCACCGCTCCCAACGCGAGAGAGAAATGTCCCACTGGTG 501

Db 61 AspArgIleLeuHisSerHisGlySerProThrArgArgLysCysAlaAsnLeuVal 80

QY 502 TCTGAGCTAACCAAGCGCTGGAGAGTGTGAGCAGAGAGAGAGAGAGAGAGAGAGAG 561

Db 81 SerGluLeuThrLysGlyTyrArgValMetGluGlnGluProThrTrpArgSerAsp 100

QY 562 AGCTGAGACACAG 621

Db 101 SerValAspThrGluAspSerGlyTyrGlyGluAlaGluGluArgProGluGlnAsp 120

QY 622 GGA 624

Db 121 Gly 121

RESULT 9

ABG59364

ID ABG59364 standard; peptide; 121 AA.

XX

AC ABG59364;

XX

DT 25-FEB-2003 (first entry)

XX

DE Human liver peptide, SEQ ID NO 38012.

XX

DE Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;

KW hypercholesterolaemia; coronary heart disease.

XX

OS Homo sapiens.

XX

PN WO200157273-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US000664.

XX

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

PD WPI; 2001-488898/53.

DR

XX

PT Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human adult liver.

PT

PS Claim 27; SEQ ID NO 38012; 658pp; English.

XX

CC The invention relates to a single exon nucleic acid probe (SEN) (1) for measuring human gene expression in a sample derived from human adult liver, comprising one of 13109 defined nucleotide sequences given in the specification (or complements/fragments). The probe hybridizes at high stringency to a nucleic acid molecule expressed in the human adult liver. (1) may be used for predicting, measuring and displaying gene expression in samples derived from human adult liver. The genes identified may be involved in genetic liver diseases such as cirrhosis, hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is associated with coronary heart disease. ABG47348-ABG59930 represent human liver single exon encoded peptides of the invention. Note: The sequence information for this patent does not appear in the printed specification

CC but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 121 AA;

Alignment Scores: 1.04e-51 Length: 121
 Pred. No.: 647.00 Matches: 121
 Score: 647.00
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 31.44% Indels: 0
 DB: 4 Gaps: 0

US-10-644-659A-1 (1-1146) x ABCS9364 (1-121)

QY 262 GGCATCGAGTACGACAAAGCCCTGAGTCTTCTCACATCAAAAGAAA 321
 |||||
 Db 1 GlyHisGlyAspGlyGlnSerSerGluValSerHisIleLysLys 20
 |||||

QY 322 GAGGTGTCCTCAAAACGGTGGTCCAGCAAGACTTACGAGAGAGGGGACGTGACCCACCTC 381
 |||||
 Db 21 GluValSerLysThrValValSerLysThrTy-GluArgGlyGlyAspValSerHisLeu 40
 |||||

QY 382 AGCCACAGGTACGAGAGGATGCTGTGCTTGAACCTGGGAGCCAGAGATGACATT 441
 |||||
 Db 41 SerHisArgTyGluArgAspAlaGlyValLeuGluProGlyGlnProGluAsnAspIle 60
 |||||

QY 442 GACAGATCTCCACAGCCACGGCTCCCAACGGGAGGAGAAAATGTGCCAACCTGGTG 501
 |||||
 Db 61 AspArgIleuHisSerHisGlySerProThrArgArgLysCysAlaAsnLeuVal 80
 |||||

QY 502 TCTGAGCTAACCAAGGGCTGGAGAGTGTATGAGCAGGAGGAGCCACATCGAGAGTGAC 561
 |||||
 Db 81 SerGluLeuThrLysGlyTyrArgValMetGluGluProThrTyrPArgSerAsp 100
 |||||

QY 562 AGCTAGACACAGGACAGCGGCTATGAGGAGAGCTGAGGAGGCGCCGAGCGAT 621
 |||||
 Db 101 SerValAspThrGluAspSerGlyTy-GlyGlyGluAlaGluArgProGluGlnAsp 120
 |||||

QY 622 GGA 624
 |||||
 Db 121 Gly 121

RESULT 10
 AAG00337
 ID AAG00337 standard; protein; 69 AA.
 AC AAG00337;
 DT 06-OCT-2000 (first entry)
 DE Human secreted protein, SEQ ID NO: 4418.
 DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.
 XX Homo sapiens.
 XX EP1033401-A2.
 XX 06-SEP-2000.
 XX 21-FEB-2000; 2000EP-00200610.
 XX 26-FEB-1999; 99US-0122487P.
 XX (GIST) GENSET.
 XX Dumas Milne Edwards J, Duclert A, Giordano J;
 PI WPI; 2000-500381/45.
 DR N-PSDB; AAC00343.
 DR
 XX

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
 XX
 PS Claim 13; SEQ ID NO 4418; 71pp + Sequence Listing; English.

CC The present sequence is a polypeptide encoded by one of a large number of
 CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
 CC prepared from total human RNAs or polyA+ RNAs derived from 30 different
 CC tissues. EST sequences usually correspond mainly to the 3' untranslated
 CC region (UTR) of the mRNA because they are often obtained from oligo-dT
 CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
 CC sequences derived from the 5' ends of mRNAs and even in those cases where
 CC longer cDNA sequences have been obtained, the full 5' UTR is rarely
 CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can
 CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
 CC are also used in diagnostic, forensic, gene therapy and chromosome
 CC mapping procedures. They are used to obtain upstream regulatory sequences
 CC and to design expression and secretion vectors

XX Sequence 69 AA;

Alignment Scores:
 Pred. No.: 4.25e-25 Length: 69
 Score: 362.00 Matches: 68
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 17.59% Indels: 0
 DB: 3 Gaps: 0

US-10-644-659A-1 (1-1146) x AAG00337 (1-69)

QY 1 ATGGTCCGGCGAAGAAAGCGGGAGGGCCAGCAAGAGGGCCCTCCGAAGATA 60
 |||||
 Db 1 MetAlaProGlyGluLysGluSerGlyGluGlyProAlaLysSerAlaLeuArgLysIle 20
 |||||

QY 61 CCACAGCCACCTCGTTCATCAGCTTGGCCGAGGTTGGCAGCAGTGGCGCAATGAGAAC 120
 |||||
 Db 21 ArgThrAlaThrLeuValIleSerLeuAlaArgGlyTyrGlnGlnTyrAlaAsnGluAsn 40
 |||||

QY 121 AGCATCAGCAGCCAGCGCTACAGCTGGCTGGCGGAGGGAGCCAGACTCACCT 180
 |||||
 Db 41 SerIleArgGlnAlaGlnGluProThrGlyTyrLeuProGlyGlyThrGlnAspSerPro 60
 |||||

QY 181 CAGCTCTTAAACCAATCACACCC 204
 |||||
 Db 61 GlnAlaProLysProIleThrPro 68
 |||||

RESULT 11
 AAY11623
 ID AAY11623 standard; protein; 70 AA.
 XX AAY11623;
 AC AAY11623;
 DT 16-JUN-1999 (first entry)
 DE Human 5' EST secreted protein SEQ ID NO:275.
 DE Human; secreted protein; EST; expressed sequence tag; diagnosis;
 KW forensic; gene therapy; chromosome mapping; signal peptide;
 KW upstream regulatory sequence; cytokine activity; cell proliferation;
 KW differentiation; haematopoiesis regulation; tissue growth regulation;
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; anti-inflammatory; tumour inhibition.
 XX Homo sapiens.
 XX WO9906439-A2.
 XX 11-FEB-1999.
 XX 31-JUL-1998; 98WO-IB001233.
 XX

PR 01-AUG-1997; 97US-00904468.
 XX (GEST) GENSET.
 XX Dumas Milne Edwards J, Duclert A, Lacroix B;
 XX WPI; 1999-153700/13.
 DR N-PSDB; AAX40341.
 XX New nucleic acids encoding human secreted proteins - obtained from cDNA
 PT libraries derived from liver, lung, large intestine, colon, thyroid and
 PT pancreas tissue.
 XX Claim 27; Page 362; 398pp; English.
 XX AAX40251 to AAX40397 represent 5' expressed sequence tags (ESTs) for
 CC human secreted proteins, and encode the proteins given in AAY11533 to
 CC AAY11679, respectively. The proteins given represent the signal peptide
 CC and an N-terminal fragment of a secreted protein. The nucleic acid
 CC sequences can be used for producing secreted human gene products. They
 CC can also be used to develop products for diagnosis and therapy. The
 CC proteins obtained may have cytokine activity, cell
 CC proliferation/differentiation activity, haematopoiesis regulating
 CC activity, tissue growth regulating activity, reproductive hormone
 CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and
 CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory
 CC activity, tumour inhibition activity or other activities. The products
 CC can be used in forensic, gene therapy and chromosome mapping procedures.
 CC The sequences can also be used for obtaining corresponding promoter
 CC sequences. The nucleic acids encoding the signal peptide can be used for
 CC directing extracellular secretion of a polypeptide or the insertion of a
 CC polypeptide into a membrane, or importing a polypeptide into a cell
 XX SQ Sequence 70 AA;

Alignment Scores:
 Pred. No.: 4,28e-25 Length: 70
 Score: 362.00 Matches: 68
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 17.59% Indels: 0
 DB: 2 Gaps: 0

US-10-644-659A-1 (1-1146) x AAY11623 (1-70)

QY 1 ATGGCTCCGGCGGAAAGGAAAGCGGGAGGCGCCAGCAAGAGCGCCCTCCGGAGATA 60
 Db 1 MetAlaProGlyGluSerGluGlyProAlaLysSerAlaLeuArgLysIle 20
 QY 61 CGCAGCCACCTGTCATCAGCTGGCCCGCGAGGTGGCAGAGTGGCGGATGAGAAC 120
 Db 21 ArgThrAlaThrLeuValIleSerLeuAlaArgGlyTrpGlnGlnTrpAlaSerGluAsn 40
 QY 121 AGCATCAGGCGAGCCCGAGGAGCTACAGGTGGCTGGCGGAGGAGCCAGCACTCACCT 180
 Db 41 SerIleArgGlnAlaGlnGluProThrGlyTrpLeuProGlyGlyThrGlnAspSerPro 60
 QY 181 CAGACTCTAAACCAATCACCC 204
 Db 61 GlnAlaProLysProIleThrPro 68

RESULT 12

ID ABB60049 standard; protein; 399 AA.
 XX AC ABB60049;
 XX DT 26-MAR-2002 (first entry)
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 6939.
 XX KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX Drosophila melanogaster.
 XX WO200171042-A2.
 XX 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US009231.
 XX 23-MAR-2000; 2000US-0191637P.
 XX 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 XX N-PSDB; ABL04152.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX Disclosure; SEQ ID NO 6939; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins (AB857737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 399 AA;

Alignment Scores:
 Pred. No.: 4,62e-15 Length: 399
 Score: 258.00 Matches: 64
 Percent Similarity: 53.29% Conservative: 25
 Best Local Similarity: 38.32% Mismatches: 50
 Query Match: 12.54% Indels: 28
 DB: 4 Gaps: 3

US-10-644-659A-1 (1-1146) x ABB60049 (1-399)

QY 718 AGCCAGTGGGCAACTTGAAGGGAGATGGCAGCAGTGGCTGATGAACACATCAATCC 777
 Db 22 SerProLeuSerSerLysValAlaMetPheAsnAsnGlnAlaThrGlnHisLysGlnSer 41
 QY 778 CAGAGCTCAATCCTTTTCAGTGAAGAGTTTGAATTACGAGCTGGCCATGTCACCCGCTA 837
 Db 42 GlnLeuLeuAsnProPheSerGlnAsp-----GlyArgAlaAlaSerProLysPro 58
 QY 838 CACAAAGGAGATGAGGGCTATGGCGCCGCCAAGAGGAAACCAAACTGCTGAAAGGCC 897
 Db 59 ThrPheSerLysAspGlnTy-GlyLysProLeuAlaGlySerLeuThrGluMetArggly 78
 QY 898 AAGCGTGTGAGGAGCATCTACAGGGAATGATGGACATGCTGCTTCATTCATTCGCACA 957
 Db 79 GlnLysAlaAsnIleHisValMetLysGluMetLeuGluLeuGlnIleleAsnSer 98
 QY 958 ATGGCTCGCCACAGACGAGATGGCAAGATCCAG-----GTTACTTTTGGAGATCTCTT 1011
 Db 99 GluGlyTyArgValLysAspGluProThrMetArgValIleProPheGlyGluLeuPhe 118
 QY 1012 GAC-----
 Db 119 AsnValSerValLeuPheThrAlaGlyIlePhePheGluLysProSerLysLeuValThr 138
 QY 1015 -----AGATACGTTCTGATTTTCAGATAAAGTAGTGGCATTCCTCATCGGTGCC 1062

CC cervical cancer. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 37 AA;

Alignment Scores:
Pred. No.: 4.38e-09 Length: 37
Score: 190.00 Matches: 36
Percent Similarity: 97.30% Conservatives: 0
Best Local Similarity: 97.30% Mismatches: 1
Query Match: 9.23% Indels: 0
DB: 4 Gaps: 0

US-10-644-659A-1 (1-1146) x AAM21625 (1-37)
QY 260 AAGGACATGGAGTGGACACAAAGCTCAGAGAAAGCCCTGAGGTTTCTCACATCAAAAAGA 319
Db 1 LysAspMetGluMetAspLysAlaGlnArgLysProLeuArgPheLeuThrSerLysArg 20
QY 320 AAGAGGTGTCCAAACCGTGTGTGTCAGCAAGACTTACGAGAGAGGAGGAGCG 370
Db 21 LysArgCysProLysArgTyrSerAlaArgLeuMetArgGluGluGlyThr 37

RESULT 15
ABB43985
ID ABB43985 standard; peptide; 37 AA.
XX
AC ABB43985;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #11491 encoded by human foetal liver single exon probe.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000669.
XX
PR 04-FEB-2000; 2000US-0180312P.
XX
PR 26-MAY-2000; 2000US-0207456P.
XX
PR 30-JUN-2000; 2000US-00608408.
XX
PR 03-AUG-2000; 2000US-00632366.
XX
PR 21-SEP-2000; 2000US-0234687P.
XX
PR 27-SEP-2000; 2000US-0236359P.
XX
PR 04-OCT-2000; 2000GB-00024263.
XX

PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human fetal liver.
XX
XX Claim 27; SEQ ID NO 36620; 639pp + Sequence Listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for measuring
XX human gene expression in a sample derived from human foetal liver. The
XX single exon nucleic acid probes may be used for predicting, measuring and
XX displaying gene expression in samples derived from human fetal liver. The
XX present sequence is a peptide encoded by a single exon nucleic acid probe
XX of the invention. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 37 AA;

Alignment Scores:
Pred. No.: 4.38e-09 Length: 37
Score: 190.00 Matches: 36
Percent Similarity: 97.30% Conservatives: 0
Best Local Similarity: 97.30% Mismatches: 1
Query Match: 9.23% Indels: 0
DB: 4 Gaps: 0

US-10-644-659A-1 (1-1146) x ABB43985 (1-37)
QY 260 AAGGACATGGAGTGGACACAAAGCTCAGAGAAAGCCCTGAGGTTTCTCACATCAAAAAGA 319
Db 1 LysAspMetGluMetAspLysAlaGlnArgLysProLeuArgPheLeuThrSerLysArg 20
QY 320 AAGAGGTGTCCAAACCGTGTGTGTCAGCAAGACTTACGAGAGAGGAGGAGCG 370
Db 21 LysArgCysProLysArgTyrSerAlaArgLeuMetArgGluGluGlyThr 37

Search completed: May 4, 2004, 08:51:36
Job time : 70 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2004, 00:04:42 ; Search time 111 Seconds
(without alignments)
5729.493 Million cell updates/sec

Title: US-10-644-659A-1

Perfect score: 1146

Sequence: 1 atggctccggcgaaagga.....tgattacgtactcaagtga 1146

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/prodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/prodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/prodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/prodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/prodata/2/ina/PCUTUS_COMB.seq.*
- 6: /cgn2_6/prodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
c 1	72	6.3	7218	1	US-08-232-463-14
c 2	39.2	3.4	289	3	US-09-007-005-17
c 3	39.2	3.4	289	3	US-09-244-796-17
c 4	38.6	3.4	5751	4	US-09-023-655-1415
c 5	38.6	3.4	11907	3	US-08-061-376-4
c 6	38.6	3.4	14255	1	US-08-320-559-1
c 7	38.6	3.4	14255	1	US-08-327-392-1
c 8	38.6	3.4	14255	1	US-08-306-691B-55
c 9	38.6	3.4	14255	3	US-08-545-860D-1
c 10	38.6	3.4	14255	5	PCT-US94-04496-1
c 11	36.2	3.2	729	3	US-09-540-236-357
c 12	36.2	3.2	1136	1	US-08-395-800A-9
c 13	36.2	3.2	1155	1	US-08-434-151-1
c 14	36.2	3.2	1155	1	US-08-208-889A-1
c 15	36.2	3.2	1155	2	US-08-433-271-1
c 16	36.2	3.2	1155	2	US-08-715-259-1
c 17	36.2	3.2	1174	5	PCT-US95-07554-3
c 18	36.2	3.2	1199	3	US-08-395-800A-5
c 19	36.2	3.2	2268	3	US-08-675-773B-4
c 20	36.2	3.2	3373	1	US-08-273-411-2
c 21	36.2	3.2	3791	3	US-08-675-773B-3
c 22	36.2	3.2	8174	1	US-07-914-281-5
c 23	36.2	3.2	8174	1	US-08-393-246-5
c 24	36.2	3.2	8174	1	US-08-525-058A-5
c 25	36.2	3.2	8174	2	US-08-696-731-5
c 26	36.2	3.2	8174	3	US-09-042-531-5
c 27	36.2	3.2	8174	5	PCT-US91-00899-3

Sequence 116, Appl
Sequence 18, Appl
Sequence 149, Appl
Sequence 2, Appl
Sequence 1, Appl
Sequence 37, Appl
Sequence 2, Appl
Sequence 1, Appl
Sequence 1702, A
Sequence 2, Appl
Sequence 3, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 11, Appl
Sequence 11, Appl
Sequence 14, Appl
Sequence 1, Appl
Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 INMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)883-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 6.3%; Score 72; DB 1; Length 7218;

```
Best Local Similarity 4.5%; Pred. No. 4.3e-11;
Matches 18; Conservative 237; Mismatches 147; Indels 0; Gaps 0;

Qy 274 GGACAAAGCTCAGAGAAGCCCTGAGGTTCTCACATCAAAAGAAAGAGGTGTCCAAA 333
Db 1438 GTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1379

Qy 334 ACGTGTGTCAGCAAGACTTACGAGAGAGGAGGACGTGACCCACCTCAGCCACAGGTAC 393
Db 1378 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1319

Qy 394 GAGAGGATGCTGCTGCTTGAACCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 453
Db 1318 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1259

Qy 454 CACAGCCAGGCTCCCAACGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 513
Db 1258 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1199

Qy 514 AAGGCTGAGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 573
Db 1198 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1139

Qy 574 GAGGACAGGCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 633
Db 1138 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1079

Qy 634 GCTGTGTCAGATCAAGCCCTTGCCTCCCGAGTAAAC 675
Db 1078 RRRRRRRRRRRATCGCAAGCTCCCTGACCTGAGCCCAAGC 1037

RESULT 2
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 625858
; GENERAL INFORMATION:
; APPLICANT: Szoostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; EARLIER FILING DATE: 1998-01-14
; EARLIER FILING DATE: 1997-01-27
; EARLIER FILING DATE: 1997-01-27
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc.feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
; US-09-007-005-17

Query Match 3.4%; Score 39.2; DB 3; Length 289;
Best Local Similarity 3.7%; Pred. No. 0.049;
Matches 6; Conservative 82; Mismatches 74; Indels 0; Gaps 0;

Qy 493 AACCTGCTGTGAGCTAACCAAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 552
Db 33 RACRARARURARACRARURGRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 92

Qy 553 AGGAGTGACAGCTGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 612
Db 93 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 152

Qy 613 GAGCAGGAGGAGTGCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 654
Db 153 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 194

RESULT 3
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szoostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; EARLIER FILING DATE: 1999-02-05
; EARLIER FILING DATE: 1997-01-27
; EARLIER FILING DATE: 1997-11-06
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc.feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
; US-09-244-796-17

Query Match 3.4%; Score 39.2; DB 3; Length 289;
Best Local Similarity 3.7%; Pred. No. 0.049;
Matches 6; Conservative 82; Mismatches 74; Indels 0; Gaps 0;

Qy 493 AACCTGCTGTGAGCTAACCAAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 552
Db 33 RACRARARURARACRARURGRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 92

Qy 553 AGGAGTGACAGCTGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 612
Db 93 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 152

Qy 613 GAGCAGGAGGAGTGCAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 654
Db 153 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 194

RESULT 4
US-09-023-655-1415/c
; Sequence 1415, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
```

SOFTWARE: PatentIn Release #1.0, Version #1.25

COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
FILING DATE: HEREWITH
CLASSIFICATION: US/09/023.655
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1415:
SEQUENCE CHARACTERISTICS:
LENGTH: 5751 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
LIBRARY: GENBANK
CLONE: G467511
US-09-023-655-1415

Query Match 3.4%; Score 38.6; DB 4; Length 5751;
Best Local Similarity 51.4%; Pred. No. 0.38; Indels 0; Gaps 0;
Matches 89; Conservative 0; Mismatches 84;

QY 465 CTCCTCCAAACGCGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAGGGCTCGAG 524
DB 428 CTCTCCCAACACGCGCCCGGAACCGGCAGGTTGTGTCGGATGGCGGCGGAGACTCGAG 369
QY 525 AGTGATGGAGCAGGAGGAGCCACCATGTGAGGAGTAGCACAGCGTAGACAGAGGACGCGG 584
DB 368 CGCGCGGTCTGAAGCCCGCGCCACCGGAGCAGGCGCGCCCTGAAGAGGCTGAGGACGA 309
QY 585 CTATGGAGGAGGCTGAGGAGGCGCGGAGGAGGCGGAGGAGTGCAGGTGGCTG 637
DB 308 TGACGAGACGAGAGCAGGCGGACGACGAGGAGGCTGCTGAGGCGCGCGCG 256

RESULT 5
US-08-061-376-4/c
Sequence 4, Application US/08061376
Patent No. 6175000
GENERAL INFORMATION:
APPLICANT: Evans, Glen A.
APPLICANT: Djabali, Malek
APPLICANT: Selleri, Lucia
APPLICANT: Parry, Pauline
TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11O23
TITLE OF INVENTION: TRANSLATION BREAKPOINT ASSOCIATED WITH ACUTE LEUKEMIAS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

QY 465 CTCCTCCAAACGCGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAGGGCTCGAG 524
DB 405 CTCTCCCAACACGCGCCCGGAACCGGCAGGTTGTGTCGGATGGCGGCGGAGACTCGAG 346
QY 525 AGTGATGGAGCAGGAGGAGCCACCATGTGAGGAGTAGCACAGCGTAGACAGAGGACGCGG 584
DB 345 CGCGCGGTCTGAAGCCCGCGCCACCGGAGCAGGCGCGCCCTGAAGAGGCTGAGGACGA 286
QY 585 CTATGGAGGAGGCTGAGGAGGCGCGGAGGAGGCGGAGGAGTGCAGGTGGCTG 637
DB 285 TGACGAGACGAGACGAGGCGGACGACGAGGAGGCTGCTGAGGCGCGCGCG 233

RESULT 6
US-08-320-559-1/c
Sequence 1, Application US/08320559
Patent No. 5633135
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
APPLICANT: Canaan, Eli
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
TITLE OF INVENTION: All-1 Region
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135ris
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/320,559
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/062,443
FILING DATE: 14 MAY 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/971,094
FILING DATE: 30-OCT-92
PRIOR APPLICATION DATA:

QY 465 CTCCTCCAAACGCGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAGGGCTCGAG 524
DB 405 CTCTCCCAACACGCGCCCGGAACCGGCAGGTTGTGTCGGATGGCGGCGGAGACTCGAG 346
QY 525 AGTGATGGAGCAGGAGGAGCCACCATGTGAGGAGTAGCACAGCGTAGACAGAGGACGCGG 584
DB 345 CGCGCGGTCTGAAGCCCGCGCCACCGGAGCAGGCGCGCCCTGAAGAGGCTGAGGACGA 286
QY 585 CTATGGAGGAGGCTGAGGAGGCGCGGAGGAGGCGGAGGAGTGCAGGTGGCTG 637
DB 285 TGACGAGACGAGACGAGGCGGACGACGAGGAGGCTGCTGAGGCGCGCGCG 233

Query Match 3.4%; Score 38.6; DB 3; Length 11907;
Best Local Similarity 51.4%; Pred. No. 0.57; Indels 0; Gaps 0;
Matches 89; Conservative 0; Mismatches 84;

QY 465 CTCCTCCAAACGCGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAGGGCTCGAG 524
DB 405 CTCTCCCAACACGCGCCCGGAACCGGCAGGTTGTGTCGGATGGCGGCGGAGACTCGAG 346
QY 525 AGTGATGGAGCAGGAGGAGCCACCATGTGAGGAGTAGCACAGCGTAGACAGAGGACGCGG 584
DB 345 CGCGCGGTCTGAAGCCCGCGCCACCGGAGCAGGCGCGCCCTGAAGAGGCTGAGGACGA 286
QY 585 CTATGGAGGAGGCTGAGGAGGCGCGGAGGAGGCGGAGGAGTGCAGGTGGCTG 637
DB 285 TGACGAGACGAGACGAGGCGGACGACGAGGAGGCTGCTGAGGCGCGCGCG 233

US-08-061-376-4

QY 465 CTCCTCCAAACGCGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAGGGCTCGAG 524
DB 405 CTCTCCCAACACGCGCCCGGAACCGGCAGGTTGTGTCGGATGGCGGCGGAGACTCGAG 346
QY 525 AGTGATGGAGCAGGAGGAGCCACCATGTGAGGAGTAGCACAGCGTAGACAGAGGACGCGG 584
DB 345 CGCGCGGTCTGAAGCCCGCGCCACCGGAGCAGGCGCGCCCTGAAGAGGCTGAGGACGA 286
QY 585 CTATGGAGGAGGCTGAGGAGGCGCGGAGGAGGCGGAGGAGTGCAGGTGGCTG 637
DB 285 TGACGAGACGAGACGAGGCGGACGACGAGGAGGCTGCTGAGGCGCGCGCG 233

US-08-061-376-4

QY 465 CTCCTCCAAACGCGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAGGGCTCGAG 524
DB 405 CTCTCCCAACACGCGCCCGGAACCGGCAGGTTGTGTCGGATGGCGGCGGAGACTCGAG 346
QY 525 AGTGATGGAGCAGGAGGAGCCACCATGTGAGGAGTAGCACAGCGTAGACAGAGGACGCGG 584
DB 345 CGCGCGGTCTGAAGCCCGCGCCACCGGAGCAGGCGCGCCCTGAAGAGGCTGAGGACGA 286
QY 585 CTATGGAGGAGGCTGAGGAGGCGCGGAGGAGGCGGAGGAGTGCAGGTGGCTG 637
DB 285 TGACGAGACGAGACGAGGCGGACGACGAGGAGGCTGCTGAGGCGCGCGCG 233

US-08-061-376-4

QY 465 CTCCTCCAAACGCGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAGGGCTCGAG 524
DB 405 CTCTCCCAACACGCGCCCGGAACCGGCAGGTTGTGTCGGATGGCGGCGGAGACTCGAG 346
QY 525 AGTGATGGAGCAGGAGGAGCCACCATGTGAGGAGTAGCACAGCGTAGACAGAGGACGCGG 584
DB 345 CGCGCGGTCTGAAGCCCGCGCCACCGGAGCAGGCGCGCCCTGAAGAGGCTGAGGACGA 286
QY 585 CTATGGAGGAGGCTGAGGAGGCGCGGAGGAGGCGGAGGAGTGCAGGTGGCTG 637
DB 285 TGACGAGACGAGACGAGGCGGACGACGAGGAGGCTGCTGAGGCGCGCGCG 233

US-08-061-376-4

QY 465 CTCCTCCAAACGCGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAGGGCTCGAG 524
DB 405 CTCTCCCAACACGCGCCCGGAACCGGCAGGTTGTGTCGGATGGCGGCGGAGACTCGAG 346
QY 525 AGTGATGGAGCAGGAGGAGCCACCATGTGAGGAGTAGCACAGCGTAGACAGAGGACGCGG 584
DB 345 CGCGCGGTCTGAAGCCCGCGCCACCGGAGCAGGCGCGCCCTGAAGAGGCTGAGGACGA 286
QY 585 CTATGGAGGAGGCTGAGGAGGCGCGGAGGAGGCGGAGGAGTGCAGGTGGCTG 637
DB 285 TGACGAGACGAGACGAGGCGGACGACGAGGAGGCTGCTGAGGCGCGCGCG 233

US-08-061-376-4

QY 465 CTCCTCCAAACGCGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAGGGCTCGAG 524
DB 405 CTCTCCCAACACGCGCCCGGAACCGGCAGGTTGTGTCGGATGGCGGCGGAGACTCGAG 346
QY 525 AGTGATGGAGCAGGAGGAGCCACCATGTGAGGAGTAGCACAGCGTAGACAGAGGACGCGG 584
DB 345 CGCGCGGTCTGAAGCCCGCGCCACCGGAGCAGGCGCGCCCTGAAGAGGCTGAGGACGA 286
QY 585 CTATGGAGGAGGCTGAGGAGGCGCGGAGGAGGCGGAGGAGTGCAGGTGGCTG 637
DB 285 TGACGAGACGAGACGAGGCGGACGACGAGGAGGCTGCTGAGGCGCGCGCG 233

US-08-061-376-4

QY 465 CTCCTCCAAACGCGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAGGGCTCGAG 524
DB 405 CTCTCCCAACACGCGCCCGGAACCGGCAGGTTGTGTCGGATGGCGGCGGAGACTCGAG 346
QY 525 AGTGATGGAGCAGGAGGAGCCACCATGTGAGGAGTAGCACAGCGTAGACAGAGGACGCGG 584
DB 345 CGCGCGGTCTGAAGCCCGCGCCACCGGAGCAGGCGCGCCCTGAAGAGGCTGAGGACGA 286
QY 585 CTATGGAGGAGGCTGAGGAGGCGCGGAGGAGGCGGAGGAGTGCAGGTGGCTG 637
DB 285 TGACGAGACGAGACGAGGCGGACGACGAGGAGGCTGCTGAGGCGCGCGCG 233

US-08-061-376-4

QY 465 CTCCTCCAAACGCGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAGGGCTCGAG 524
DB 405 CTCTCCCAACACGCGCCCGGAACCGGCAGGTTGTGTCGGATGGCGGCGGAGACTCGAG 346
QY 525 AGTGATGGAGCAGGAGGAGCCACCATGTGAGGAGTAGCACAGCGTAGACAGAGGACGCGG 584
DB 345 CGCGCGGTCTGAAGCCCGCGCCACCGGAGCAGGCGCGCCCTGAAGAGGCTGAGGACGA 286
QY 585 CTATGGAGGAGGCTGAGGAGGCGCGGAGGAG

RESULT 7
US-08-327-392-1/c
Sequence 1, Application US/08327392
Patent No. 5633136
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
APPLICANT: Canani, Eli
TITLE OF INVENTION: ALL-1 Polymers
TITLE OF INVENTION: Antibodies
TITLE OF INVENTION: Treatment
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn
STREET: One Liberty Place - 4
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 IN
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/327
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/971, 09
FILING DATE: 30-OCT-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/888, 83
FILING DATE: 27-MAY-92
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/805, 09

RESULT 8
US-08-306-691B-55/c
; Sequence 55, Application US/08306691B
; Patent NO. 5734039
; GENERAL INFORMATION:
; APPLICANT: Calabretta, Bruno
; APPLICANT: Skorski, Tomasz
; TITLE OF INVENTION: ANTISENSE
; TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seidel, Gonda, Lavoragna & Monaco, P.C.
; STREET: Two Penn Center, suite 1800
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/306,691B
; FILING DATE: September 15, 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A.
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; TELEX: NO. 5734039e
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 14255 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-306-691B-55

```

Query Match 3.4%; Score 38.6; DB 1; Length 14255;
Best Local Similarity 51.4%; Pred. No. 0.63;
Matches 89; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

45	CTCCCCAACGGGAGGAGAAAATGTGCCA	CTCGTGTCTGAGCTTAACCAAGGCGTGGAG	524
Qy			
219	CTCCCCAACCGGCCCGGACGGTTCGCCGAT	GCGGCGCGAGACCTTGCAG	160
Db			
525	AGTGATGGAGCAGGAGGAGCCCAATGAGGAG	GTGACAGCGTATAGACACAGAGACACACGG	584
Qy			
159	CGCCCGCTCGAAGCCCGGGCCCAACCGAG	CAGGGCGGCGCTTAAGAGAGTCGAGACGA	100
Db			

Qy 585 CTATGGAGGAGGCTGAGGAGAGGCCCGACGAGTGGAGTGCAGGTGGCTG 637

Db 99 TGACGAAGACCAAGACGAGGCGGACGACGAGGAGGCTGCTGAGCGGCGGCGG 47

```

RESULT 9
US-08-545-860D-1/c
; Sequence 1, Application US/08545860D
; Patent No. 6040140
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaani, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: For Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:

```

```

, APPLICATION NUMBER: US 07/805,093
, FILING DATE: 11-DEC-1991
, ATTORNEY/AGENT INFORMATION:
, NAME: Deluca Esq., Mark
, REGISTRATION NUMBER: 33,229
, REFERENCE/POCKET NUMBER: TJU-1262
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: (215) 568-3100
, TELEFAX: (215) 568-3439
, INFORMATION FOR SEQ ID NO: 1:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 14255
, TYPE: nucleic acid
, STRANDEDNESS: single
, TOPOLOGY: linear
, ANTI-SENSE: No
, US-08-545-860D-1

```

	Query Match	3.4%	Score 38.6;	DB 3;	Length 14255;
	Best Local Similarity	51.4%;	Prod. No. 0.63;		
	Matches	89;	Conservative	0; Mismatches	84; Indels
				0; Gaps	0;
QY	45	CTCCCCAACGGGAGGAGAAAATGTC	CCAACTGGTGTCTAGCTAACCAAGGGCTGGAG	524	
Db	219	CTCCCCAACACGCCCGGAAACCGG	CGCAGGTGTGTGCGATGCGCGGCGGAGACCTGCGAG	160	
QY	525	AGTGATGGAGCAGGAGGAGCCCAT	ATGGAGGAGTGACACAGCTAGACACAGAGGACACGCGG	584	
Db	159	GCGCGCTGAGGCCCGGCCACCGG	AGAGCAGGCGCGCCCTGAAGAGGCTGAGGACGA	100	
QY	585	CTATGGAGGAGGCTGAGGAGAGG	CCCCGAGCAGGATGGATGTCAGTGGCTG	637	
Db	99	TGACGAGACGAGCAGGCGCGAC	ACACAGAGAGGCTGTGAGGGCGGCGCGG	47	

```

RESULT 10
PCT-US94-04496-1/c
; Sequence 1, Application PC/TUS9404496
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaani, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESSEE: Norris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04496
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: DeLuca Bag., Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-1242
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14255
; TYPE: nucleic acid

```

STRANDEDNESS: single
TOPOLOGY: linear
ANTI-SENSE: NO
PCT-US94-04496-1

Query Match 3.4%; Score 38.6; DB 5; Length 14255;
Best Local Similarity 51.4%; Pred. No. 0.63; Indels 0; Gaps 0;
Matches 89; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 465 CTCCTCCCAACGCGGAGGAGAAATGTGCCAACCTGTGTGTGAGCTAACCAAGGCTGGAG 524
DB 219 CTCCTCCCAACACGCGCGCGGAAACCGCGAGGTTGTGTGCCATGCGCGCGGAGCTGCAG 160
QY 525 AGTGATGACAG 584
DB 159 CGCGCGCTGCAAGCG 100
QY 585 CTATGAGGAG 637
DB 99 TGACGAG 47

RESULT 11
US-09-540-236-357/c
Sequence 357, Application US/09540236
Patent No. 6673910
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
FILE REFERENCE: 2709.2005-001
CURRENT APPLICATION NUMBER: US/09/540.236
CURRENT FILING DATE: 2000-04-04
NUMBER OF SEQ ID NOS: 3840
SEQ ID NO 357
LENGTH: 729
TYPE: DNA
ORGANISM: M. catarrhalis
US-09-540-236-357

Query Match 3.2%; Score 36.2; DB 4; Length 729;
Best Local Similarity 57.5%; Pred. No. 0.64; Indels 0; Gaps 0;
Matches 65; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
QY 425 ACCGAGAGATGACATGACAGATGACAGATGACAGATGACAGATGACAGATGACAGATGAC 484
DB 695 AGCAGCGCTATCTCATTCATCAAAAAATCCACATCCAGCCATCATACGCAAGCAAAA 636
QY 485 AATGTCCTCAACCTGTGTCTGAGCTAACCAAGGCTGGAGAGTGTGAGAGAGAGAGAGAG 537
DB 635 ATGTGCTCCACCTCATCTTGACGGTAAGCAGATCTTGGTCTTGAGCGATAAG 583

RESULT 12
US-08-395-800A-9/c
Sequence 9, Application US/08395800A
Patent No. 5807732
GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B
APPLICANT: LENNON, GREGORY
APPLICANT: ROQUIER, SYLVIE
APPLICANT: GIORGI, DOMINIQUE
APPLICANT: KELLY, ROBERT J
TITLE OF INVENTION: GDP-L-FUCOSE: BETA-D-GALACTOSIDE
TITLE OF INVENTION: 2-ALPHA-L-FUCOSYLTRANSFERASES DNA SEQUENCES ENCODING THE
TITLE OF INVENTION: SAME, METHOD FOR PRODUCING THE SAME AND A METHOD OF
TITLE OF INVENTION: GENOTYPING A PERSON
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: ORLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VIRGINIA

COUNTRY: USA
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/395,800A
FILING DATE: 28-FEB-1995
CLASSIFICATION: 435
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1136 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 39..1133
US-08-395-800A-9

Query Match 3.2%; Score 36.2; DB 1; Length 1136;
Best Local Similarity 50.9%; Pred. No. 0.82; Indels 0; Gaps 0;
Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY 472 ACCGCGGAGAGAGAAATGTCCAAACCTGTGTGTGAGCTAACCAAGGCTGGAGAGTGTG 531
DB 215 ACCCGGCGGAG 156
QY 532 GAGCAGGAG 591
DB 155 CGACAGGCTAGTGGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 96
QY 592 GGAGAGGCTGAGGAG 640
DB 95 GAGGACAGAGACTAGCAGGAG 47

RESULT 13
US-08-434-151-1/c
Sequence 1, Application US/08434151
Patent No. 5700671
GENERAL INFORMATION:
APPLICANT: Prieto, P.
APPLICANT: Smith, D.
APPLICANT: Cummings, R.
APPLICANT: Koppick, J.
APPLICANT: Mukerji, P.
APPLICANT: Moremen, K.
APPLICANT: Pierce, J.
TITLE OF INVENTION: Transgenic Animals Producing Oligosaccharides and Glycocon
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Donald O. Nickey
ADDRESSEE: ROSS Products Division
ADDRESSEE: Abbott Laboratories
STREET: 625 Cleveland Avenue
CITY: Columbus
STATE: Ohio
COUNTRY: United States
ZIP: 43215
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh System 7.1
SOFTWARE: ClarisWorks 1.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/434,151

FILING DATE: 02-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/209,132
FILING DATE: 09-MAR-1994
TELECOMMUNICATION INFORMATION:
TELEPHONE: (614) 624-7080
TELEFAX: (614) 624-3074
TELEX: No. 5700671e
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1155 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Single
TOPOLOGY: Unknown
MOLECULE TYPE: Cloned cDNA representing the product of a human
MOLECULE TYPE: genomic DNA segment;
DESCRIPTION: GDP-L-fucose--D-galactoside 2-alpha-fucosyl-
HYPOTHETICAL: transferase
ANTI-SENSE:
FRAGMENT TYPE: Entire amino acid sequence provided.
ORIGINAL SOURCE: Human Epidermal Carcinoma Cell line
ORGANISM:
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE: Human Epidermal Carcinoma Cell line
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 19
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD: DNA sequencing and restriction analysis
OTHER INFORMATION: The encoded product of nucleotide SEQ ID NO: 1: is the enzyme
PUBLICATION INFORMATION:
AUTHORS: V.P. Rajan et al
TITLE:
JOURNAL: Journal of Biological Chemistry
VOLUME: 264
ISSUE:
PAGES: 11158 - 11167
DATE: 1989
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:

US-08-434-151-1

Query Match 3.2%; Score 36.2; DB 1; Length 1155;
Best Local Similarity 50.9%; Pred. No. 0.83;
Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY 472 ACGCGGAGGAGAAATGTGCCAAGCTGGTCTCAGCTAACCAAGGGCTGGAGAGTGATG 531
Db 214 ACCCGGAGGAGGATGGCCATCTGGGGGTGTCCACGAGGGGGTCTGGACACAGGAT 155
QY 532 GAGCAGGAGGAGCCACATGAGGAGGTGACAGCGTACACAGAGGACAGCGCTATGGA 591
Db 154 GCACAGGCGCTAGGCGCATGTGGAAGCTGCTTTGATGATATGAGGAGGAAGATTACAGA 95
QY 592 GGAGAGGCTGAGGAGAGGCCCGGAGCAGGATGGAGTGCAGGTGGCTGTGG 640
Db 94 GAGGACACAGACTAGCAGGAAGGCCAGGACAGAGCTGACGATGGCTCCGG 46

RESULT 14
US-08-208-889A-1/c
Sequence 1, Application US/08208889A
Patent No. 5750176
GENERAL INFORMATION:
APPLICANT: Prieto, P., Smith, D., Cummings, R., Kopchick, J., Mukerji, P.,
APPLICANT: Moremen, K., Pierce, J.
TITLE OF INVENTION: Transgenic Production of Oligosaccharides and
TITLE OF INVENTION: Glycoconjugates
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Donald O. Nickey
ADDRESSEE: ROSS Products Division
ADDRESSEE: Abbott Laboratories
STREET: 625 Cleveland Avenue
CITY: Columbus
STATE: Ohio
COUNTRY: United States of America
ZIP: 43215
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS Version 6.21
SOFTWARE: WordPerfect Version 6.0a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/208,889A
FILING DATE: 09-MAR-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA: No. 5750176 applicable
TELECOMMUNICATION INFORMATION:
TELEPHONE: (614) 624-7080
TELEFAX: (614) 624-3074
TELEX: No. 5750176e
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1155 base pairs
TYPE: Nucleic acid
STRANDEDNESS: Single
TOPOLOGY: Unknown
MOLECULE TYPE: Cloned cDNA representing the product of a
MOLECULE TYPE: human genomic DNA segment
DESCRIPTION: GDP-L-fucose- -D-galactoside 2-alpha-fucosyl-transferase
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE: Entire amino acid sequence provided.
ORIGINAL SOURCE: Human Epidermal Carcinoma Cell line
ORGANISM:
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE: Human Epidermal Carcinoma Cell line
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 19
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD: DNA sequencing and restriction analysis
OTHER INFORMATION: The encoded product of nucleotide SEQ ID NO:
OTHER INFORMATION: 1: is the enzyme, GDP-L-fucose- -D-galactoside 2-alpha-fuc
OTHER INFORMATION: which has the amino acid sequence described in SEQ. ID NO:
OTHER INFORMATION: is responsible for the synthesis of 2'-fucosyllactose.
PUBLICATION INFORMATION:
AUTHORS: V.P. Rajan, et al.

Db 134 nGlyGluAlaaspGluProSerPheSerAspProGluAspPheValAspValSer-- 153
Qy 378 CCTCAGCCAGGTACAGAGGGATGCTGGTGTGCTTGAACCTGGGAGCCAGAGATGA 437
Db 154 -----GluGluGluLeuLeuGlyAspValLeuLeuAspArgProGlnGluAlaAs 170
Qy 438 C---ATTGACAGAATCTCCACAGCCAGGC-----TCCCAACGGGGAGGAG 482
Db 170 pGlyLeuAspSerValLeuValValAspAsnValProGlnValGlyProAspArgLeuG 190
Qy 483 AAAATGTGCCAACCTGGTGTGAGCTA----- 510
Db 190 uLeuLeuAsnValLeuHisLysIlePheSerLysPheGlyLysIleLeuAsnAspTy 210
Qy 511 -----ACCAAGGGCTGAGAGTGTGAGCAGCAGGAGGAGCCAC 548
Db 210 rTy-ProGluGluAspGlyLysThrLysGlyTyIlePheLeuGluLysAlaSerProAl 230
Qy 549 ATGGAGGAGTACAGCGTAGACACAGAGGAGCAGCGGTATGAGGAGAGGCTGAGGAGAG 608
Db 230 aHisAlaValAspAlaValLysAsnAlaAsp---GlyTy----- 242
Qy 609 CCCCAGCAGCAGTGGAGTGCAGGTGGTGTGCTCAGGATCAAGCGCCCTTGCCCTCCCA 668
Db 243 -----LysLeuAspLysGlnHisThrPheAr 251
Qy 669 GGTAAACAGATTACA-----CAGAACTCACTGCA 701
Db 251 gValAsnLeuPheThrAspPheAspLysTyMetThrIleSerAspGluTrpAspIlePr 271
Qy 702 AGCCCAACAGAAATATAGCCAGTGGCACTTGAAGGAGATGGCAGCAGTGGCTGA 761
Db 271 oGluLysGlnProPheLysAspLeuGlyAsnLeuArg-----TyIleLeuG 287
Qy 762 TGAACACATCAATCCCAAGAGCTCAATCCTTTTCACTGAAGAGTTTATTACAGAGTGGC 821
Db 287 uGlu-----AlaGluCysArgAspGlnTy-SerValIlePheGlu----- 300
Qy 822 CATGTCCACCGCCTACAAAGAGGATGAGGGC-----TATGGCCGCCCAAGA 872
Db 301 -----SerGlyAspArgThrSerIlePheTrpAsnAspValLysAs 314
Qy 873 AGGAACCAAACTGCTGAAGGGCCCAAGCGTGTGCTGAGGAGCAGCATCTACAGGAAATGAT 932
Db 314 pProValSerIleGluGluArgAlaArgTrpThrGlu----- 326
Qy 933 GGACATGTCTTATTATCTGCACATAGTGGCTGCGCCACAGAGATGGCAGATCCAGGT 992
Db 327 -----ThrTyValArgTrpSerProLysGlyThrTyIleLeuAl 339
Qy 993 TACTTTT-----GGAGATCTCTTTGACATACGTTG 1025
Db 339 aThrPheHisGlnArgGlyIleAlaLeuTrpGlyAspLysPheLysGlnIleGlnAr 359
Qy 1026 TATTTGATAAAGTAGTGGGATCTCATCGTGCCAGGAAACATGACTGGTAGACTT 1085
Db 359 gPheSerHisGlnGlyVal-----GlnLeuIleAspPh 370
Qy 1086 T 1086
Db 370 e 370

RESULT 6

E2WTHW

glutenin, high molecular weight chain precursor - wheat

C:Species: Triticum aestivum (common wheat)

C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 16-Jul-1999

C:Accession: A24107

R:Sugiyama, T.; Rafalski, A.; Peterson, D.; Soll, D.

Nucleic Acids Res. 13, 8729-8737, 1985

A:Title: A wheat HMW glutenin subunit gene reveals a highly repeated structure.

A:Reference number: A24107; MUID:86093674; PMID:3001648

A:Accession: A24107
A:Molecule type: DNA
A:Residues: 1-838 <SUC>
A:Cross-references: GB:X03346; NID:G21784; PIDN:CAA27052.1; PID:9736319
A:Experimental source: cv. Yamhill
C:Comment: Glutinins, like gliadins, are high in glutamine and proline but differ in
C:Superfamily: glutenin
C:Keywords: seed; storage protein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-838/Product: glutenin, HMW chain #status predicted <MAT>

Alignment Scores:
Pred. No.: 0.213 Length: 838
Score: 123.00 Matches: 75
Percent Similarity: 34.90% Conservative: 29
Best Local Similarity: 25.17% Mismatches: 98
Query Match: 5.98% Indels: 96
DB: 1 Gaps: 13

US-10-644-659A-1 (1-1146) x E2WTHW (1-838)

Qy 9 GGGCGAAAAGAAAGCGGGAGGCCCGCCAGCCAGAGCGCCCTCCGGAAGATACGCACAGC 68
Db 464 GlyGlnLysGlyGlnGlnProGlyGlnGlnProGlyGlnGlnGlnGlnGln 483
Qy 69 CACCTGTGTATCATCAGCTTGGCCGAGGTGGCAGCAGTGGCGCAATGAGAAACAGCATCAG 128
Db 484 GlnProGly-----GlnGlnGln 489
Qy 129 GCAGGCCAGGAGCCT-----ACAGCTGGCTGCGGGAGGACCCAGGACTC 176
Db 490 -GlnGlyGlnGlnProGlyGlnGlyGlnProGlyTyTyTyProThrSerProGlnGlnSe 509
Qy 177 ACCTCAGCTCTTAAACA-----ATCAC 200
Db 509 rGlyGlnGlyGlnGlnProGlyGlnTrpGlnGlnProGlyGlnGlnProGlyTyTyTy 529
Qy 201 ACCCTTACTTCCACACCAGAAAGCTCAGAGTGCCTCCAAAGTGCACCCCGCTGCCAGA 260
Db 529 rProThrSerProLeuGlnProGlyGlnGlyGlnProGlyTyTyTyTyTyTyTyTy 549
Qy 261 AGACATGAGATGACAAAGCTCAGAAAGCCCTGAGGTTTCTCATCATCAAAAGAA 320
Db 549 nGlnProGlyGlnGlyGlnGln-----Pro--GlyGlnLeuGlnGlnProAla 564
Qy 321 AGAGGTGTCCAA-----AACGTGTGTCCAGCAAGACTTACGAGAG 359
Db 565 GlnGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnVal 584
Qy 360 AGAGGGAGCTGAGCCACCTCAGCCACAGGTA-----CGAGAG 398
Db 585 GlnGlnGlnGlnProAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 604
Qy 399 GGATGCTGTGTCTTGAACCTGGGCGAGCCAGAGATGACATTGACAGAACTCTCCACAG 458
Db 605 Gly-----GlnGlnProGlyGln 610
Qy 459 CCAGCTCTCCCAACCGGAGGAGAAATGTGCCAACCTGGTGTCTGAGCTAACAGAGG 518
Db 611 GlyGlnGlnProAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 624
Qy 519 CTGGAGATGATGGACAGGAGGCCACATGGAGGAGTGCAGCGGTAGACACAGAGAGA 578
Db 625 -----GlnGlyGlnGlnGlnGlnProGlyGlnGlnGlnGlnGlnGln 641
Qy 579 CAGCGGCTATGG-----AGAGAGGCTGAGGAGAGCCCGGAGCAGAGA 620
Db 642 GlnProTrpTy 661
Qy 621 TGGAGTGCAGGTGCG-----TGTGTGTCCAGGATCAA----- 650
Db 662 TrpGlnGlnProGlyGlnTrpGlnGlnProGlyGlnGlnGlnProGlyTyTyTyTyTy 681

Db 3309 ProLaserThrHisTyrTyrSerAspSerAspTyrArgHisGlyAlaArgAlaAspLys 3328
 QY 844 GGAGATGAGGCTATGCGCGCCCAAGAAAGAACCAAACTGCTGAAGCGGCAAGCGT 903
 Db 3329 TyrGlyProGlyProMetGlyProLysHisProSerLysAsnLeuAlaProAlaAla--- 3347
 QY 904 GCTGAGGACACATCTACAGGAAATGATGACATGTGCTTCATTATCTGCACAATGGCT 963
 Db 3348 -----IleSerSerLysArgSer 3353
 QY 964 GCCACAGACGATGCGCAAGATCCAG 990
 Db 3354 LysHisArgLysGlnGlyMetGluGln 3362
 RESULT 11
 S24457
 Hypothetical protein R087.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 09-Sep-1997
 C:Accession: S24457; S41036
 R:Sulston, J.; Du, Z.; Thomas, K.; Wilson, R.; Hillier, L.; Staden, R.; Halloran, N.; Gw
 awkins, T.; Ainscough, R.; Waterston, R.
 Nature 356, 37-41, 1992
 A:Title: The C. elegans genome sequencing project: a beginning.
 A:Reference number: S24457; MUID:92168156; PMID:1538779
 A:Accession: S24457
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-458 <SUL>
 A:Cross-references: ENBL:Z12017; NID:96812; PID:96813
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1992
 C:Genetics:
 A:Introns: 35/2; 145/1; 170/1; 428/2
 Alignment Scores:
 Pred. No.: 0.366 Length: 458
 Score: 119.50 Matches: 89
 Percent Similarity: 38.36% Conservative: 56
 Best Local Similarity: 23.54% Mismatches: 161
 Query Match: 5.81% Indels: 72
 DB: 2 Gaps: 19
 US-10-644-659A-1 (1-1146) x S24457 (1-458)
 QY 115 GAGAACATCAGCAGCGCCAGAGGCTACAGCTGGCTGCGGAGGCCAGCAG--- 171
 Db 68 LysGlnSerValLysAlaLysValHisGlyPheLysGlnThrPheAlaGluVal 87
 QY 172 GACTCAGCTCAGCT-----CCTAAACCAATCACACCCCTACTTCA----- 213
 Db 88 AspAlaProLysValLysProGluProLysSerProAspAsnSerProProArgGly 107
 QY 214 -----CACAGAAAGCTCAGAGTCCCGCCAAAG----- 240
 Db 108 LysArgGlnArgHisAspSerAspAsnSerProProArgProSerArgLysArgAsnAsp 127
 QY 241 -----TGCCACCCCGCTGCCAGAGGACATGGATGGACAAAGCTCAGAGAA 291
 Db 128 SerAspAsnSerProProArgProSerArgAsnArgHisAspSerAspLysAspAsnSer 147
 QY 292 GCCCTGAGTTTCTCATCATCAAAAGAGGTGTCCAAAACGGTGGTCAGCAAGACT 351
 Db 148 ProPro-ArgArgArgHisAspSerAspAsnSerProProArgProSerArgLys-I 167
 QY 352 TACGACAGAGG-----AGGGACGTGACCCACCTCAGCCAGGATGACAGAGG 399
 Db 167 LeArgGluGluSerProSerAlaArgAsnArgArgSerPro---ProArgThrArg-Arg 185
 QY 400 GATGTGTGTGTTGAACCTGGGCGAGCCAGAGAT-----GACATT 441
 Db 186 AspArgHisAspSerAspAsnSerProProArgAsnArgSerArgArgAspSerAspAsn 205
 QY 442 GACAGATCTCTCCACAGCCACCGCTCCCAAGCGGAGGAGAAATGTGCCAACCTGGTG 501

Db 206 SerProProArgArgArgProSerSerProAlaArgArgLysAspAspLeu--- 224
 QY 502 TCTGAGCTAACCAAGGGGTGGAGATGATGGAG-----CAGAGAGAGCCC 546
 Db 225 SerProProArgLysSerArgLysIleGluGluProLysLysIleLysLysGluGluPro 244
 QY 547 ACATGGAGGATGACAGCGTAGACACAGAGGACAGCGGCTATGGAGGAGCGCTGAGGAG 606
 Db 245 -----AspSer---AspThrGluThrSerGlyArgThrLeuGluGly----- 257
 QY 607 AGGCCCGAGCAGGATGAGTGGCTGCTGTCAGGATCAAGCGCCCTTGGCCCTCC 666
 Db 258 -----LysArgSerGlyLeuGlnSerAla-----ArgAspLeuLysGlu 270
 QY 667 CAGGTAACAGATTACAGAGAACTCAACTCCAAAGCCCAACACAGAAATATAGCCAGTG 726
 Db 271 Glu-----SerAspLysLeuArgAlaLysAsnSerLysMetPheGluGluMet 286
 QY 727 GGC---AACTTGAAGGGAGATGGCAGCAGTGGGCTGATGAACACATCAATCCAGAG 783
 Db 287 AspThrSerValSerGlyArgPheAlaAspThrValTyrArgGlnLysGlnThrLysLys 306
 QY 784 CTCATCTCTTCAGTGAAGAGTTTATACAGAGCTGGCCATGCTCCACC---CGCTACAC 840
 Db 307 LysGlyLysAspSerGluGluAspGlnAlaLysLysGluArgGluThrLysLysThrGlu 326
 QY 841 AAAGGAGATGAGGGGTATGGCGCCCAAGAAAGAACCAAACTGCTGAAGGGCCAG 900
 Db 327 GluLeuLysGluLysTyrLysSerTrpAsnLysGlyValAlaGlnIleGluAspArgArg 346
 QY 901 CGTGTGAGGACACATCTACAGGAAATGATGGACATGTGCTTCATTATCTGCACAATG 960
 Db 347 AlaGlnLeuGluGluMetAlaArgValAlaAlaGlu-----ProMet 360
 QY 961 GCTCGCCACAGCAGATGCCAAGATCCAGGTACTTTTGGAGATCTCTTTCACAGATAC 1020
 Db 361 AlaArgAlaArgAspAspAlaMetAsnAlaHisLeuLysGluVal----- 376
 QY 1021 GTTCGTATTTCAGATAAGTATGTCATTCATCGTCGCGCAGGAAACATGACTGCTGTA 1080
 Db 377 LeuHisAlaAlaAspProMetAlaAsnMetIleGlnLysLysLysArgAspThrAlaIle 396
 QY 1081 GACTTTGAAGGAGAGATGCTATGGCAAGCGCCAGATGACCATGCTGTGTG 1128
 Db 397 Asp---ArgGlyGluLeuValTyrProSerTyrHisGlyHisPheVal 411
 RESULT 12
 B30843
 glutenin high molecular weight chain Ax2 precursor - wheat
 C:Species: Triticum aestivum (common wheat)
 C:Date: 19-May-1989 #sequence_revision 19-May-1989 #text_change 15-Nov-1996
 C:Accession: B30843
 R:Anderson, O.D.; Greene, F.C.
 submitted to GenBank, January 1989
 A:Reference number: A94515
 A:Accession: B30843
 A:Molecule type: DNA
 A:Residues: 1-815 <BND>
 C:Superfamily: glutenin
 C:Keywords: seed; storage protein
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-815/Product: glutenin Ax2 chain #status predicted <MAT>
 F:108-773/Region: glutamine/glycine/proline-rich
 Alignment Scores:
 Pred. No.: 0.464 Length: 815
 Score: 118.50 Matches: 90
 Percent Similarity: 33.97% Conservative: 34
 Best Local Similarity: 24.66% Mismatches: 131
 Query Match: 5.76% Indels: 111
 DB: 2 Gaps: 18


```

516 GlnGlnPro-----GlnGlnGlnGlnProGlyTyTyProThrSerPro 530
621 -----TGAGTGCAGGTGGCTGTGTGTCAGGAT 647
531 GlnGlnSerGlyGlnGlnGlnLeuGlnGlnTrpGlnGlnSerGlyGlnGln--- 549
648 CRAAGCCCGCTCCCTCCAGGTAAACAGATTACAGAGAACTCAACTGCAA----- 701
550 -----ProGlyHis-----TyProThrSerProLeuGlnProGly 561
702 -----AGCCACAGAAATATAGCCAGTGGGCACTT 734
562 GlnGlnProGlyTyTyProThrSerProGlnGlnIle-----GlyGlnGly 578
735 GAAAGGGAGATGGCAGCGTGGCTGATGAACATACATCCAGAGTCAATCCTTT 794
579 Gln----- 579
795 CAGTGAAGATTGATTACAGAGTGGCATGTCACCCGCTTACAAAGGAGATGAGG 854
580 -----GlnProGlyGlnLeuGlnGlnProThrGlnGlnGlnGly 593
855 CTATGGCGCCCAAA-----AGAAGAACCAAAAC 884
594 GlnGlnProGlyGlnGlnGlnGlnGlnGlnProGlyGlnGlnGlnGlnGln 613
885 TGCTGAAGGGCCCAAGCGTCTGAGGAGCACATCTACAGGAAATGATGACATGTG--- 941
614 ProGlyGlnGlnGln-----GlnProGlyGlnGlnProGlyTyTyPro 629
942 -----CTTCATTATCTGCACATGGCTGCCACAGAGAGTGGCAAGATCCAGT 992
630 ThrSerLeuGlnGlnSerGlyGlnGlnGlnGlnProGlyGlnTrpGlnGlnProGly 648

RESULT 14
S74544
translation initiation factor IF-2 - Synchocystis sp. (strain PCC 6803)
N:Alternate names: hypothetical protein elr0744
C:Species: Synchocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 02-Feb-2001
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis
s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S74544
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1001 <KAN>
A:Cross-references: EMBL:D90900; GB:AB001339; NID:gl651768; PIDN:BRAL6696.1; PID:gl651768
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: infB
C:Superfamily: translation initiation factor IF-2; translation elongation factor Tu homolog
C:Keywords: GTP binding; nucleotide binding; P-loop
F:496-609/Domain: translation elongation factor Tu homology <ETU>
F:502-509/Region: nucleotide-binding motif A (P-loop)
F:606-609/Region: nucleotide-binding motif A (P-loop)
F:642-644/Region: GTP-binding NXXD motif
F:642-644/Region: GTP-binding SAK/L motif
F:508,509,529,606,607,609,642/Binding site: Mg-GTP (Lys, Thr, Thr, Asp, Ser) #

Alignment Scores:
Pred. No.: 0.672 Length: 1001
Score: 116.50 Matches: 81
Percent Similarity: 35.61% Conservative: 39
Best Local Similarity: 22.69% Mismatches: 111
Query Match: 5.66% Indels: 126
DB: 2 Gaps: 18
```

```

US-10-644-659A-1 (1-1146) x S74544 (1-1001)
QY 12 CGAAAGCAAGAGCGGGAGGCGCCAGCAAGAGCGC-----CCTCCGAGATACG 62
Db 89 LysGlnGlnLysSerGlyProSerProAlaArgProThrProProAlaArgProLys 108
QY 63 CACAGCCACCTGGTCATCAGCTTGGCCCGAGG---TTGGCAGCAGTGGCGGAATGAGAA 119
Db 109 LeuGlnAlaProLysAlaProThrProGlnProProValAlaLys-AlaSerAlaPr 128
QY 120 CAGCATCAGGCGGCCAGGAGCTTACAGCTGGCTCGCGGAGGAGCCAGGACTCACC 179
Db 128 OlySileGlnLysGlnGlnGluPro-----AlaG1 138
QY 180 TCAAGCTCTTAAACAATCACACCCCTACT-----TC 212
Db 138 nGluAlaProLysSerValAlaProProThrGlnProLeuAlaProProValProSe 158
QY 213 ACACGAGAAAGCTCAGAGTGCCTCCAAAGTCGCCA----- 246
Db 158 kLeuGlnSerProProSerLysProAlaProProThrProProAlaLysLysAlaAlaPr 178
QY 247 -----CCCCGCTGCAGAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCTGAGGT 302
Db 178 oAlaProArgLeuAlaGlyProProGlyArgThrAlaSerProAsnLysThrAlaValPr 198
QY 303 TTCACATCAAAAAGAAAGAGGTGTCCAAAACGGTGTCTCAGCAAGACTTACGAGAGAGG 362
Db 198 oAlaProAlaLysProLysValAsnArgProGluLysValSerLeuLysAspAsnArgL 218
QY 363 A-----GGGAGCGTGAGCCACTCAGCCACAGCTCAGCCACAGAGGAGGTGCT-- 405
Db 218 yGlnAlaArgSerProGlyAsp-----ArgGluGluLysValAlaAl 232
QY 406 -GGTGTGCTTGAACCTGGCGGAGGAGAGATGACATTGACAGAAATCTCCACAGCCACGG 464
Db 232 eAlaAlaProGluProProLysProLysValGlnLeuArgArg-----Pr 247
QY 465 CTCGCCAAGCGGAGGAGGAGAAATGTCCAACTGTGTCTGAGCTTAACAAAGGCTGAGG 524
Db 247 oLysProProArgProGluGluAspGluAsnLeuProGluLeuLeu----- 262
QY 525 AGTGATGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 584
Db 263 -----GluPheProProLeuSerArgGlyLysGlyValAspGlyAspAsnArgPAL 279
QY 585 CTATGAGGAGAG-----GCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 638
Db 279 aAspAspGlyAspLeuLeuSerThrGluLysProLysPro----- 292
QY 639 GGTGAGGATCAAGCGCCCTTCCCTCCAGGTAACAGATTACAGAGAACTCAACTG 698
Db 293 -----LysLeuLysArgProThrProArgLeu----- 302
QY 699 CAAAGCCCAACAGAAATATAGCCAGTGGGCAACTTGAAGAGGAGATGCGACAGTGGGC 758
Db 303 -----GlyLysProAspGlnTrpG1 309
QY 759 TGATGAACACATACAAATCCAGAACTCAAT----- 789
Db 309 uAspAsp-----GluAspGluLysAlaAsnLysAlaLysAlaAsnLysGlyLysAr 327
QY 790 -----CCTTTTCAGTGAGAGGTTTATTACGAGTGGCTGAGTCCAGCCGCTACACAAAGG 845
Db 327 sArgProLysMetAspAspAspAspAspAspAspAspAspAspAspAspAspAspAsp 342
QY 846 AGATGAGGCG-----TATGGCGGCCCC----- 867
Db 342 YAspAsnGlyProLysProThrLeuValSerLeuSerLeuSerLeuAlaArgProLysProLys 362
QY 868 -----AAGAGAGGAGCAAAACTGCTGAAGGAGGAGGAGGAGGAGGAGGAGGAGG 903
Db 362 sSerLeuAlaAlaLysProSerThrProThrValAlaLysValLysLys 378
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: May 4, 2004, 08:49:09 ; Search time 20 Seconds
(without alignments)
5916.336 Million cell updates/sec

Title: US-10-644-659A-1
Perfect score: 2058
Sequence: 1 atgctcgggcaaaagga.....tgattacgctactcaagtga 1146

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues
Total number of hits satisfying chosen parameters: 778829

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlp
-MODEL=frame_n2p.model -DBUS=xlp
-Q/cgn2_1/USPTO.spool.p/US10644659/runat_04052004_084903_28566/app.query.fasta_1.1287
-DB-issued_Patents_AA_QFWT=fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10644659 @CGN 1 1 27 @runat_04052004_084903_28566 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCUTS.COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB ID	Description
1	138.5	6.7	748	4	US-09-252-991A-18427	Sequence 18427, A
2	134.5	6.5	454	4	US-09-252-991A-30164	Sequence 30164, A
3	134	6.5	416	4	US-09-252-991A-30923	Sequence 30923, A
4	133.5	6.5	1225	4	US-09-252-991A-25018	Sequence 25018, A
5	130.5	6.3	442	4	US-09-252-991A-31848	Sequence 31848, A
6	130.5	6.3	1739	4	US-09-976-594-76	Sequence 76, Appl
7	130	6.3	586	4	US-09-252-991A-28601	Sequence 28601, A
8	127.5	6.2	499	4	US-09-252-991A-18439	Sequence 18439, A
9	127	6.2	638	4	US-09-252-991A-27068	Sequence 27068, A
10	126.5	6.1	609	4	US-09-252-991A-23771	Sequence 23771, A
11	125	6.1	741	4	US-09-252-991A-27062	Sequence 27062, A
12	124.5	6.0	256	4	US-09-252-991A-25404	Sequence 25404, A

13	124.5	6.0	482	4	US-09-252-991A-17580	Sequence 17580, A
14	124	6.0	1113	4	US-09-252-991A-29215	Sequence 29215, A
15	121.5	5.9	437	4	US-09-252-991A-26942	Sequence 26942, A
16	121	5.9	422	4	US-09-252-991A-27632	Sequence 27632, A
17	121	5.9	1182	4	US-09-287-354-6	Sequence 6, Appl
18	120.5	5.9	728	4	US-09-252-991A-23613	Sequence 23613, A
19	120	5.8	507	4	US-09-252-991A-28026	Sequence 28026, A
20	119.5	5.8	479	4	US-09-252-991A-32884	Sequence 32884, A
21	119.5	5.8	409	4	US-09-252-991A-23598	Sequence 23598, A
22	118.5	5.8	508	4	US-09-252-991A-25549	Sequence 25549, A
23	118.5	5.8	529	4	US-09-252-991A-18245	Sequence 18245, A
24	118	5.7	407	4	US-09-252-991A-31517	Sequence 31517, A
25	118	5.7	863	4	US-09-252-991A-26099	Sequence 26099, A
26	117.5	5.7	429	4	US-09-252-991A-33116	Sequence 33116, A
27	117.5	5.6	524	4	US-09-198-452A-52	Sequence 52, Appl
28	117.5	5.7	528	4	US-09-252-991A-19913	Sequence 19913, A
29	117.5	5.7	822	4	US-09-252-991A-22479	Sequence 22479, A
30	117	5.7	373	4	US-09-252-991A-29008	Sequence 29008, A
31	117	5.7	379	4	US-09-252-991A-28428	Sequence 28428, A
32	117	5.7	625	4	US-09-252-991A-19871	Sequence 19871, A
33	117	5.7	788	4	US-09-252-991A-28171	Sequence 28171, A
34	117	5.7	989	4	US-09-252-991A-17435	Sequence 17435, A
35	116.5	5.7	593	4	US-09-252-991A-31061	Sequence 31061, A
36	116.5	5.7	693	4	US-09-252-991A-24059	Sequence 24059, A
37	116	5.6	411	4	US-09-252-991A-25491	Sequence 25491, A
38	116	5.6	414	4	US-09-252-991A-31110	Sequence 31110, A
39	116	5.6	842	4	US-09-252-991A-21753	Sequence 21753, A
40	116	5.6	822	4	US-09-252-991A-21920	Sequence 21920, A
41	115.5	5.6	277	4	US-09-252-991A-25033	Sequence 25033, A
42	115.5	5.6	300	4	US-09-252-991A-27253	Sequence 27253, A
43	115.5	5.5	449	4	US-09-252-991A-23908	Sequence 23908, A
44	115.5	5.6	564	4	US-09-252-991A-28765	Sequence 28765, A
45	115	5.6	273	4	US-09-252-991A-23651	Sequence 23651, A

ALIGNMENTS

RESULT 1
US-09-252-991A-18427
; Sequence 18427, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18427
; LENGTH: 748
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18427

Alignment Scores:	3.53e-05	Length:	748
Pred. No.:	138.50	Matches:	97
Score:	30.92%	Conservative:	27
Best Local Similarity:	24.19%	Mismatches:	111
Query Match:	6.73%	Indels:	166
DB:	4	Gaps:	21

US-10-644-659A-1 (1-1146) x US-09-252-991A-18427 (1-748)

QY 23 GCGGGAGGGCCACCCAGACGCCCTCCGGA-----AGATAGCAGCAGCCACC 73
DB 88 AlaGlyThrAlaArgProArgAsnArgGlySerArgProArgProAlaGlnPro---- 106

QY 74 TGCTCATCAGCTTGGCCCGAGGTTGGC-----AGCAGTGGCGAATGAGACACATCA 127
Db 107 -----ValGlyGlnArgSerProGlySerArgArgThrArgPro 119
QY 128 GGCAGGCCAGGACCTACAGGCTGGCTGCGCGGAGGACCCAGGACTCACCTCAAGCTC 187
Db 120 HisArgProArgHis---SerGlyGlnCysArgPheAlaProGlyProGln----- 135
QY 188 CTAACCAATCACACCCCTACTTCAC-----ACC 217
Db 136 ---ArgProGlyHisProGlyGluHisGlnGlnHisGlnArgGlySerSerProThr 154
QY 218 AGAAAGCTCAGAGTGGCCCAAGTGCACCCCGCTGCCAGA----- 260
Db 155 GlnGlnProGlnArgProGlyArgArgHis-ProGlnProArgHisGlnGlnHisArgHi 174
QY 261 ----AGACATGAGATGGCAAAAGCTC----- 284
Db 174 sProArgArgTrpArgTyrArgProLeuGlyAlaArgLeuGlnArgProArgIleGlyAs 194
QY 285 -----AGAGA 289
Db 194 pAlaThrValArgArgHisAlaProValHisArgHisGlyHisGlyGluLeuProGlnAr 214
QY 290 AAGCCCTGAGTTTTCATCATCAAAAGAA----- 320
Db 214 gSerValAspGlyArgAlaHisArgArgAspProArgProGlyLeuGlyAlaValArgAr 234
QY 321 -----AGAGTGTCCAAACGGTGGTCAGCAAGACTTACGAGAGAGAGGGG 367
Db 234 gArgArgHisArgArgGlyAspGlnArgGlyAlaGluGluAlaPheArg---ArgArgAs 253
QY 368 ACGTGACCACTCAGCCACAGGTACGAGAGGGATGTGTGTGCTTGAACCTGGGCAGC 427
Db 253 pProGlnProProAlaProArgLeuArg-----LeuLe 264
QY 428 CAGAGATGACATTGACAGATCTCTCCA----- 455
Db 264 uArgGln-----ProProValGlyProGlyGlnArgArgLeuAlaAspAr 279
QY 456 -CAGCCAGGCTCCCAACGGCGAGGAG----- 482
Db 279 gGlnProGlnLeuProAlaGlnProGluProAlaAlaGluProArgLeuAspArgProTr 299
QY 483 -AAATGTGCCAAGTGTGTCTGAGCTTAACCAAGGCTGGAGAGTGTGAGCAGAGG 541
Db 299 pArgLeuAlaGlnProGly-----HisGlnArgGlyAlaAlaLe 312
QY 542 AGCCCAATGAGGAGTGCACAGCGTAGA-----CACAGAGCAGCGCGTATGGAG 592
Db 312 uAlaGly-----GlnArgArgSerGlyLeuHisProArgProArgLeuTrpAr 328
QY 593 GAGAGCTGAGGAGAGGCCGAGCAGGATGAGTGCAGGTGGCTGTGTCAGGATCAAGC 652
Db 328 sProGly---AlaAspGluArgLeuArgHisProAlaGlyArgArgGlnValProGlnAl 347
QY 653 GCGCCCTGCGCTCCAGGTAAACAGATTACAGAAACTCACTCAAGAAACCCCAACAGA 712
Db 347 sProAlaArgGluGluLeuGlnArgAlaGlnArgAlaAlaLeuGlnArgProValAs 367
QY 713 A-----ATATAGCCAGTGGGCAACTTGAAGGGAGATGGCAGCAGTGGG 757
Db 367 pProAspGlnArgLeuGluProPheArgGlnArgHisArgGluGlnProValVal-- 386
QY 758 CTGATGAACACATACATCCCAAGAGCTCAATCCTTCAGTGAAGAGTTGATTACGAGC 817
Db 387 -----LeuHisGlnGlyPro----- 391
QY 818 TGGCCATGTCTCCACCGCTTACAAAGGAGATGAGGGCTATGGCCGCCGCCCAAGAGGAA 877
Db 392 -----ProProLeuAlaGlnAr 397
QY 878 CCAAAACTGCTGAAAGGGCCAGCGTGTGAGGAGCAGCATCTACAGGGAAATGATGGACA 937

Db 397 gArgAsnLeuArgMetGlyArgProAlaArgAlaValAlaGlnGlyLeuProAlaHi 417
QY 938 T 938
Db 417 s 417
RESULT 2
US-09-252-991A-30164
; Sequence 30164, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30164
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30164
Alignment Scores:
Pred. No.: 7,046-05 Length: 454
Score: 134.50 Matches: 85
Percent Similarity: 35.47% Conservative: 48
Best Local Similarity: 22.67% Mismatches: 147
Query Match: 6.54% Indels: 96
DB: 4 Gaps: 15
US-10-644-659A-1 (1-1146) x US-09-252-991A-30164 (1-454)
QY 11 GCGAAAGAAAGCGGGAGGCGCCAGCAAGAGCGCTCCGGAGATAGCA----- 64
Db 5 AlaProArgProAlaGly-----ProAlaProAlaGlyArgGlyAlaProGly 20
QY 65 -----CAGCCACCTGGTTCATCAGCTTGGCCCGAGTGGCAGCAGTGGCGGA---ATG 115
Db 21 GluGlySerProGlyTrpSerProAlaProGlyGlu-----ThrAlaGlyArgCysThr 38
QY 116 AGACAGCATCAGGAGGCGCCAGGAGCTACAGGCTGGCTGCCGGAGGAGCCAG-GAC 174
Db 39 ArgArgAlaAlaGly-----AlaAlaCysArgAspAlaProGlySer 52
QY 175 TCACCTCAAGCTCCTAAACCAATCACACCCCTACTTCCACCCAGAAAGCTCAGAGTGCC 234
Db 53 SerArgLysSerProSerAlaAlaProProAlaThrGlySerThrArgTrpArgGlnPhe 72
QY 235 CCA---AAGTCCGACCCCGCTGCCAGAGGA----- 264
Db 73 ProCysSerSerProProGlyLeuAlaGlyGlyArgAlaAlaSerSerProTrpAla 92
QY 265 -----CATGGA----- 270
Db 93 AlaArgSerSerArgArgSerTrpArgLeuGlnHisGlyThrGlnGlnAlaGlnArgVal 112
QY 271 GATGGCAAAAGCTCAGAGAAAGCCCTGAGGTTTCTCATCATCAAAAGAAAGAGGTGCC 330
Db 113 AspGlyGluGlnGlnGlyIleValGlyValLeuHisGluAspAlaGlnValGluHisGln 132
QY 331 AAAACCGTGTCTCAGCAAGACTTACGAGAGAGGAGGAGCTGAGGCACCTCAGCCACAGG 390
Db 133 GlnGlnAlaGlyHisGlnAlaLeuAspArgHisGlyGluAspGluHisValGlnArgArg 152
QY 391 TACGAG---AGGATGCTGTGTGTGCTTGAACCTGGCGAGCCAGAGATGAC----- 438

Db 153 GlyluAlaAlaAspGlnArgGlnAlaGluLeuGlyGluHisGlyAspGlnArg 172
 QY 439 -----ATTGACAGAACTCTCCACAGCCAGGGTCCCAACGCGGAGGAGAAATGTGCC 492
 Db 173 ArgGlyGluGlnArgAlaGluHisGlnHisLeuArgGluLeuArgValGluGluCysHis 192
 QY 493 AACTGTGTCTGAGCTAAACCAAGGCTGGAGAGTATGAGCAGAGGAGCCCATGG 552
 Db 193 ProLeuAlaAlaGluLeu-----ArgGluGluThrAlaHis 204
 QY 553 AGGAGTGACAGCGTAGACACAGAGCAGCGGCTATGGAGAGAGCGCTGAGAGAGGCC 612
 Db 205 ArgProThrGlyIleAlaProGluGlnArgAlaAspGlnArgGluGluAlaAlaGly 224
 QY 613 GAGCAGGATGAGTGCAGGTGCTGTCAGGATCAAGCGCCCTTGCCCTCCAGGTA 672
 Db 225 GluGluGlnGlyAspGln-----ArgLeuArgGlnGlnProGlyGlyAla 240
 QY 673 AAC-----AGATTTACAGAGAAATC 693
 Db 241 GlyLeuAlaAlaGlnAlaGlyValGluGlyArgGlyGluCysArgAlaThrLeuValVal 260
 QY 694 AACTGCAAGCCCAACAGAAATATAGCCAGTGGCAACTTGAAGGGAGATGGCAGAG 753
 Db 261 AspAspGluAlaGlyArgPheHisGlyProGluGlyAsnGlyGlyAspGluAlaAspGlu 280
 QY 754 TGGCTGATGATCAACATACATATCCAGAGCTCAATCTTTCAGTGAAGAGTTTGATTAC 813
 Db 281 GlnAlaGluGlnHisPheLeuAla-----AspGlnGlnGlyAspVal 294
 QY 814 GAGCTGGCCATGTCACCCGCTACACAAAGGAGATGAGGCTATGGCGCCCAAGAA 873
 Db 295 GlnArgAlaGluGlyArgGlnAlaAlaGlyAspAspArgValGlyGluGlnArgGln 314
 QY 874 GGAACCAAACTGCTGAAAGGCCCAAGCGTCTGAGGAGCAGATCTACAGGGAATGATG 933
 Db 315 ArgThrGlyHisAlaValAlaHisGlnAlaAlaGlnGlyLeu----- 329
 QY 934 GACATGTCTTCATTATCTGCACATGCTCGCCACAGCAGAT 978
 Db 330 -----LeuSerGlnHisArgGlnAsp 336

RESULT 3

US-09-252-991A-30923
 ; Sequence 30923, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Mario J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 30923
 ; LENGTH: 416
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-30923

Alignment Scores:
 Pred. No.: 7,59e-05 Length: 416
 Score: 134.00 Matches: 111
 Percent Similarity: 34.87% Conservative: 40
 Best Local Similarity: 25.64% Mismatches: 140
 Query Match: 6.51% Indels: 142
 DB: 4 Gaps: 27

US-10-644-659A-1 (1-1146) x US-09-252-991A-30923 (1-416)

QY 8 CGGCGGAAAGAAAGCGGGGAGGGCCACCAAGAGCGCCCTCCGAGATACCCA--- 64
 Db 24 ArgAlaArgArgThrGlyGlyAlaGlyLeuAlaArgArgProGlyArgGlyAlaAla 43
 QY 65 -----CAGCCACCCCTGTCATCAGCTTCGCGGAGGTGGCAGAGTGGGGAATGAGA 118
 Db 44 ThrCysArgLeuProGlyGlnProArgTyrProArgThrGly---ProGlyArg--- 60
 QY 119 ACAGCATCAGCGAGCCCGAGGCTCAGGCTGCGCGGAGGAGCCAGGACTCAC 178
 Db 61 -----GlyArgProArgSerHisArgValGlyThrGlyGluArgGlnArgArgGly 77
 QY 179 CTCAGGCTCTAAACCAATCACACCCCTACTTCACCCAGAAAGCTCAGA--- 229
 Db 78 LeuArg-----ArgProValLeuValArgArgGlyArgArgAlaGlyGly 92
 QY 230 -----GTGCCCCAAGTGCACCCCGCTGCCAGAGGACATGAGATGACAAAGCT 283
 Db 93 GlyAlaValSerProGluGlnTyrProGly-----ArgAspValGlyLeuProGlnArg 110
 QY 284 CAGAGAAGCCCTGAGGTTTCTCACATCAAAAGAAAGAGGTGTCCAAAC--- 335
 Db 111 AlaLeuArgAlaLeuArgLeuGlyCysGlyGluAspArgArgHisPro-GlnLeuProAla 130
 QY 336 -----GTTGTCAGCAAGACTTACAGAGAGGAGCGACCTGAGCCACCTCAGCCACA 388
 Db 130 aglyAlaAlaGlyGlyAlaAspLeuArgAlaAlaArg---ArgHisProProGlyProAla 149
 QY 389 GGTA-----CGAGAGGATGCTGGTGTCTTGAACCTCGGCAGCCAGAGATGACATG 442
 Db 149 aLeuAlaArgAlaGly-AlaAlaMet-----GlyGlnProGlnArg----- 163
 QY 443 ACAGAACTCTCCACAGCCCGCTCCCAACCGGAGGAGAAATGTGCAACCTCGTG- 501
 Db 164 -----ArgProGlyAspGlyArgArgArgArgCysAlaGlyArgValP 178
 QY 502 -----TCTGAGCTAAACCAAGCGCTGGAGAGTGTATGAGCAGAGGAGCCACAT 550
 Db 178 roGlyArgArgGlyAspLeu-ProArg-----ArgAlaAlaPro--- 190
 QY 551 GGAGGAGTACAGCGTAGACACAGAGCAGCGGCTATGAGAGAGAGCGGTGAGGAGAGGC 610
 Db 191 -----AlaAlaAlaProArgGlnArgProAlaAla-----AlaGly 202
 QY 611 CCGAGCAGGATGAGTGCAGGTGCTGCTCAGGATCAAGCGCCCTTGCCTCCAGG 670
 Db 203 ProArgHisLeu-----SerGlyProGlyAlaGlyThrProGlyArg 216
 QY 671 TAAACAGATTACAGAGAACTCACTGCAAGCCCAACAGAAATATATAGCCAGTGGCA 730
 Db 217 -----AlaValLeuAlaAlaGlyAlaGlyArgThrLeu----- 227
 QY 731 ACTTGAAGGAGATGCGACAGTGGCTGATGAACACACATACAAT-----CCAGAGAGC 784
 Db 228 -----GlyHisGlyValAlaAlaCysLeuLeuThrGluGlyCysAlaArgArg 244
 QY 785 TCAATCCTTCAGTGAAGAGTTGATGATGAGCTGCGCATGTCCACCCCTCACAAG 844
 Db 245 ArgAlaAlaThrThrPheArgThrArgThrGlyTyrThrCysProProAlaSerArgPro 264
 QY 845 ---GAGATGAGGCTATGGCGCCCAAGAGAACCAAACTGTGTAAGGCGCCAGC 901
 Db 265 GlyArgPheArgCysArgArgSerProSerArg-----ProGlyProAla 279
 QY 902 GTGCTGAGGAGCATCTTACAGGGAATGATGACATGTGCTTATTATTCGACATGG 961
 Db 280 CysArgArgAlaThrThrAlaThrLysProLeuProAlaSerThrThrAlaPro-AsnG 299
 QY 962 CTGCCCACACAGCAGATGG-----CAAGATCCAGG 991
 Db 299 y-----TyrCysArgThrProAlaAlaArgArgSerArgProCysGlnGlyCysG 317

QY 992 TTACTTTGGAGATCTCTTTGACAGATACGTTGCTGATTTCAGATATAAGTAGTGGGATTC 1051
Db 317 Y-----GlyAlaGlyGlyse 322
QY 1052 TCATGCGTGCAGGA-----ACATG 1072
Db 322 rAlaSerCysArgAlaMetProSerThrProThrProGlyAlaArgProGlyThrTr 342
QY 1073 GACTGGTAGACTTTGAGGAGAGATGCTATGCGCAAGGCCGAGA-----1115
Db 342 pCysGlyArg-----SerAlaArgProArgSerProIleSerTh 356
QY 1116 ----TGACATGTTGTGATTAGCTACTCAA 1142
Db 356 rAlaSerThrCysCysTrpTrpSerThrArg 366
RESULT 4
US-09-252-991A-25018
; Sequence 25018, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25018
; LENGTH: 1225
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25018
Alignment Scores:
Pred. No.: 0.000143 Length: 1225
Score: 133.50 Matches: 100
Percent Similarity: 32.00% Conservative: 36
Best Local Similarity: 23.53% Mismatches: 156
Query Match: 6.49% Indels: 133
DB: 4 Gaps: 18
US-10-644-659A-1 (1-1146) x US-09-252-991A-25018 (1-1225)
QY 12 CGAAAGGAAAGCGGGAGGGCCACCCAGAGCGCCCTCCGGAAGATACGCACAGCCAC 71
Db 30 ArgValGlyGlyArgGlyGlyProArgHisGlnGlnArgArgAspProAlaLeuArgHis 49
QY 72 CTGGTGCATCAGCTTGGCCGAGGTTGGCAGCAGTGGCGGAATGAGACAGCATCAGGCA 131
Db 50 ProGlyArgArgThrGlyProGlyThrAlaGluHisSerProGluSerHisPro-As 69
QY 132 GCGCCAGGAGCCCTACA---GGTGGCTGCGGGAGGAGCCAGGAC-----174
Db 69 pArgLeuArgProAlaProGlyGlyGlyGlyProAlaGlyThrGlyArgProGlyLeuGln 89
QY 175 -----TCACCTCAAGC 185
Db 89 nArgArgAlaGlyAlaAlaAlaGlyLeuAlaGlyGlnArgProGlyIleHisProGlnAr 109
QY 186 TCCTAAACCAATCACACCCCTACTTCACACAGAAAGCTCAGAGTGGCCCAAGTCGCC 245
Db 109 gArgArgArgAlaGlnProAlaThrArgHisGlyAlaGlyCysArgAlaAlaGlyGlyG 129
QY 246 ACCCGCCTCCAGAGGACATGGATGGA-----CAAG 281
Db 129 uAlaArgArgAspProGlyHisGlyGlyGlyHisArgHisProGlnGlyArgCysGln-G 149
QY 282 CTCAGAGAAAGCCCTGAGGTTTCTCACATCAAAAAGAAAGAGGTGTCAAAACGGTGT 341

Db 149 lyArgGlnProGlyThrAspArgCysGlyGlnArgArgAlaAlaAlaAlaAlaAlaAla 169
QY 342 CAGCAAGACTTACGAGAGAGGAGGACGTCAGCCACCTCAGCCACAGGTACGAGAGGA 401
Db 169 lArgArgProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 183
QY 402 TCGTGGTGGCTTGAACCTGGCCAGCAGAGATGACATTGACAGAAATCCCTCCACAGCCA 461
Db 184 ----ValAlaProGlyProGlyLeuGlnArgPheGlyAlaAspGlnArgAlaValAla 202
QY 462 CGGCTCCCAACGCGGAGGAGAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTG 521
Db 202 laArgProGlu-----SerGlnProArgAlaG 211
QY 522 GAGAGTGAT-----530
Db 211 ly-AlaGlyGluArgGlnAlaProSerArgAspArgGlnHisGlnCysGlnArgAla 230
QY 531 -----GGAGCAGGAGGAGCCACATGGAGGAGTGACAGCGGTAGACAC 572
Db 231 AlaGluArgLeuHisArgGlyArgProGlyTyHisProAlaAlaAlaAlaAlaAla 250
QY 573 AGAGGACAGCGGTATGAGGAGAGGCTGAGGAGGAGCCGAGGAGGAGGAGGAGGAGGAG 632
Db 251 ArgGlyAlaAla-----ArgArgArgLeuGlyAlaValArgLeuGlyArgHisArgArg 268
QY 633 GGC-----TGTGTCAGGATCAAGCGCCCTTCCCTCCAGGTAA 674
Db 269 GlyAspGlnArgHisProGlnValGlyArgProArgArgProAlaArgLeuArgGlnArg 288
QY 675 CAGATT-----TACAGAGAACTCACTGCAAGCCCAACAGAAATATAGCCAGTGGG 728
Db 289 SerLeuArgArgTrpArgProArgProGluArgArgGlyGlyGluTrpArgProAlaPhe 308
QY 729 CAACTTGAAGGAGAGTGGCAGCAGTGGCTGATGAACACATACATCAATCCAGAGTCAA 788
Db 309 ArgArgGluArg-----PheProGlnProGln 317
QY 789 TCCTTTTCAGTGAAGAGTTTATTACAGAGTGGCCATGTCCACCG-----833
Db 318 -----ArgAlaValProArgThrGlyAlaAspHisProArgHisArgPro 333
QY 834 -----CCTACAAAGGAGATGAGGGCTATGGCGGCC-----866
Db 334 AlaTyArgAlaLeuArgGlnProValHisArgArgProPheAlaAlaProGlyAla 353
QY 867 ---CAAAGAAGGAACCAAACTCTGAAAGGGCCAGCGTGTCTGAGGAG-----CACAT 917
Db 354 GlyLeuGlnArgArgGly-AlaGlyArgGlnArgArgAlaValGlnLeuArgHisLe 373
QY 918 CTACAGGGAATG-----ATGGACATGTCTTCATTATC-----TGCAC 956
Db 373 uTyProSerLeuGlyValLeuGlyAlaAspLeuProValAlaValProGlyAlaAla 393
QY 957 AATG-----GTCGCCACAGCAGATGGCAAGATCCAGGTACTTT 998
Db 393 rLeuSerGluArgLeuHisProAlaHisHisGlnArgArgGlyArgLeu-----Ph 410
QY 999 TGGAGAT 1005
Db 410 eAlaAsp 412

RESULT 5
US-09-252-991A-31848
; Sequence 31848, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A

Qy	635	CTGTGTCAGGATCAAGCGCCCTTGCCTCCCAAGGTAACAGAGATTTACAGAGAAACTCA	694
Db	334	---TrpLysAlaSerSerAla-----ArgSerArgGlyThr	344
Qy	695	ACTGCAGACCCCAACAGAAATATAGCCCAAGTGGGCAACTTGAAAGGGAGATGCCAGCACT	754
Db	345	AlaTrpLysPro-----CysSerAspProAlaHis	354
Qy	755	GGGTGATGAACACATACATCCACAGAGTCATCCCTTCAGTGAAGAGTTTGATTCAG	814
Db	355	GlyValGlnHisAlaargValProArgCysPheValThrIleProAlaProAlaAlaGlu	374
Qy	815	AGCTGGCCATGTCCCGCCCTACACAAAGGAGATGAGGGCTATGGCGGCCCAAGAG	874
Db	375	AlaGlyProAlaPheProThrLeuGlyValValIargargAlaAlaCys-----Arg	391
Qy	875	GAACCAAACTGCTGAAGGGCCA	898
Db	392	ArgProArgArgLeuAlaGlyPro	399

RESULT 8

```

RESULI 8
US-09-252-991A-18439
; Sequence 18439, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18439
; LENGTH: 499
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18439

```

Alignment Scores:		
Pred. No.:	0.000374	499
Score:	127.50	109
Percent Similarity:	30.10%	Conservative: 43
Best Local Similarity:	21.58%	Mismatches: 162
Best Match:	6.20%	Indels: 191
Query Match:	4	Gaps: 23
DE:		

US-10-644-659A-1 (1-1146) X US-09-252-991A-18439 (1-499)

QY	23	GCGGGAGGCCCCAGACCCAGAGGGCCTCCGGAAGATACGCACGCCACCTTGGTCATCA	82
Db	3	AlaAlaThrThrThrProSerAlaProLys---ArgTyrSerArgProAspTrpProThr	21
			:::
QY	83	GCTTGCCCGAGGTGGC-----AGCAGTGGG-----	109
Db	22	ArgTTPArgSerAlaGlyArgSerSerProThrProThrTrpSerSerGlySerValSer	41
QY	109	-----	109
Db	42	AlaArgArgTTPProArgProThrArgArgGlySerThrValalaThrLeuProLaThr	61
QY	110	-----CGAATGAACAAGCATCAGCAGCACCCCAGGAGCTA-----	145
Db	62	ProIleThrArgLeuSerAlaSerProLeuArgProAspSerLeuProSerHisAlaSer	81
QY	146	-----CAGGCTGGCTGGCG-----GAGGGA	166
Db	82	PheProGlyAlaAlaThrAlaAlaLeuAlaGluGlusSerCysArgAlaIleAlaAspPro	101
			:::
Ov	167	CCGAGGACTCACCTCAAGCTCTAAACCACATCACACGCCCTACTCTCACACCAGAAAAGCTC	226

432 --ArgHisArgProArgArg-----GlnLeuGlyAspArgArg 445
 QY 1022 TTGCTATTTCAGATAAGTAGTGGCATTCTCATGCTGCCAGAAACATGGAGTGTAG 1081
 Db 445 lnA-g-----ProGlyAlaAlaArgArgGlyArgGalaGlyLeuProG 460
 QY 1082 ACTTTGAAGGA 1092
 Db 460 lyGlyAspGly 463

RESULT 9
 US-09-252-991A-27068
 ; Sequence 27068, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 27068
 ; LENGTH: 638
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-27068

Alignment Scores:
 Pred. No.: 0.000474 Length: 638
 Score: 127.00 Matches: 79
 Percent Similarity: 33.23% Conservative: 27
 Best Local Similarity: 24.72% Mismatches: 123
 Query Match: 6.17% Indels: 90
 DB: 4 Gaps: 13

US-10-644-659A-1 (1-1146) x US-09-252-991A-27068 (1-638)

QY 11 GCGAAAGGAAGACGGGGAGGCCAG-----CCAAGA 43
 Db 53 AlaleuArgArgThrGlyArgAlaGlnHisProProValProAlaGlnArgArgArg 72
 QY 44 GCGCCCTCGGAAGATACGACAGCCACCCCTGGTCTATCAGCTGGCCGAGGTGGCAGC 103
 Db 73 AlaProAlaGlyArgAlaAlaSerPro----- 81
 QY 104 AGTGGCGGAATGAGAACAGCATCAGCCAGGCCAGGAGCTACAGGCTGGCTGCCGGAG 163
 Db 82 -----GlylysProArgArgGlyLeuAlaGlyAlaThrAla 93
 QY 164 GACCCAGGACTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 223
 Db 94 GlyProArg-----ProAspProArgAlaAlaProArgGly 105
 QY 224 CTCAGAGTCCCAAGTCCGACCCGCTCCGACAGGAGCATGAGATGACACAAAGCT 283
 Db 106 ValAlaArgIleGlnArgArgProAlaArgArg-----GlnAla 120
 QY 284 CAGAGAACCCCTGAGGTTCTCATCAATAAAGAAAGAGGTGTCCAAACGGTGGTCA 343
 Db 121 ArgArgArgProAlaGlyProAlaAlaThrGlyArgAlaArg---ProHisArgGlnPro 139
 QY 344 GCAAGACTTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 403
 Db 140 GlyThrMetValArgLeuProGlyArgArgGlyGlyAlaAlaProGlyArgArg----- 157
 QY 404 CTGGTGTGCTGAACCTGGGCGCAGAGATGACATTGACAAATCTCCACAGCCCA-- 461
 Db 158 -----LeuProHisProAlaAlaAlaArgGlnProAlaProAsp-ProProGlnProG 175

QY 462 -----CGCTCCCAAC-----CGCAGAGAGAAATGTGCCAAC 496
 Db 175 yProArgGlnAlaGlyGlyLysArgGlnProGlyProGlyArgArgArgAlaArgI 195
 QY 497 TGGTCTGAGTAAACAGGGCTGGAGAGTGTGAGCAGAGGAGGAGGAGGAGGAGGAGG 556
 Db 195 eglyGlnProAlaArgAlaGlyArgAlaAspArgGlyLeuAlaAlaArgAla-- 214
 QY 557 GTGACAGCGTACACAGAGGAGCAGCGGCTATGGAGGAGA-----GGCTGAGGAGAGGC 610
 Db 215 -----ArgArgAlaThrLeuGlyGlnArgLeuArgHisArgGlyAlaGlyArg 233
 QY 611 CCGACAGGATGAGTGCAGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 670
 Db 233 gArgArgGlyTpaArg----- 238
 QY 671 TAAACAGATTACACAGAACTCAACTGCAAAAGCCCAACAGAAATATAGCCAGTGGGCA 730
 Db 239 -ArgGlnCysArgArgProArgArg-GlyAlaAlaAlaArgThrArgGlnProAlaAla 258
 QY 731 ACTTGAAGGAGATGGCAGCAGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 790
 Db 258 hrLeuSerGlyArg-----ValAlaProHisProArgProAspArgAlaSerP 274
 QY 791 CTTTCAGTGAAGAGTTGTATTACGAGCTGGCCATCTCCACCGCCTACACAAAGGAGATG 850
 Db 274 ro-----AlaThrValArgArgHisArgArgGlnT 284
 QY 851 AGGGCTATGCGCGCCCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGCT 903
 Db 284 hrGlyArgGlyAlaProGlyProGlyHisProGlyThrAlaGlyAlaArgArg 301

RESULT 10
 US-09-252-991A-23771
 ; Sequence 23771, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 23771
 ; LENGTH: 609
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-23771

Alignment Scores:
 Pred. No.: 0.00052 Length: 609
 Score: 126.50 Matches: 110
 Percent Similarity: 32.66% Conservative: 35
 Best Local Similarity: 24.77% Mismatches: 144
 Query Match: 6.15% Indels: 155
 DB: 4 Gaps: 24

US-10-644-659A-1 (1-1146) x US-09-252-991A-23771 (1-609)

QY 3 GGCTCCGGCGGCAAGAAAGCGGGAGGCGCCAGCAGCAGCGCCCTCCGGAAGATACG 62
 Db 217 GlyHisGlyArgThrGlyArgHisGly-----ArgAlaThrGluArgThr 231
 QY 63 CAC-----AGCCACCTGTATCATAGCTTGGCCCGAGGTGGCAGCA----- 104
 Db 232 HisProAspProTyrProGlyHisArgTyrArgSerArgProSerAlaAspLeuAspArg 251

105 GTGGCGAATAGAACAGCATCAGGAGGCGCCAGGAGCCCTACAGGCTGCTCCCGGAGG 164
Db
252 ArgGlyArgGlyAlaLeuSerArgProGlyProGly-----LeuGlyArgGlyAla 268
Qy
165 GACCAGACTCAGCTCAAGCTCCTAAACCAATCACACCCCTTCTACACAGCAAGG 224
Db
269 AspProGly-----AlaGlyArg 274
Qy
225 TCAGAGTGCCTCAAGTCGCGCCCGCTGCCAGAGGACATGGAGATGGCAAGCTC 284
Db
275 GlyArgCysProGlyArgAlaLeuProAspGlnArgPro-----GlyArgLeu 290
Qy
285 AGAGAAAGCCCTGAGTCTTCCATCATCAAAAGAGAGGTGTCAAAACCGTGTGCTAG 344
Db
291 ArgAlaLeuProProGlyGlyCysArgGlnAlaProGlyThrAlaAlaArgCysGlyAla 310
Qy
345 CAAGACTTAGCAGAGAGGAGGAGCGTGCAGCCACCTCAGCCACAGTACGAGGAGTGC 404
Db
311 AspArgLeuArgGlyLeu----- 316
Qy
405 TGGTGTCTTGAACCTGGGCGAGCCAGAGATGATGATGATGATGATGATGATGATGAT 464
Db
317 -----GlnArgProLeuGlnArgArg 323
Qy
465 CTCCTCC----- 470
Db
324 LeuProAlaArgArgGlyGlyAlaLeuArgAlaGlyHisAlaGlyProAlaProProGly 343
Qy
471 AACCGGAGAGAAATGTGCCAACCTGGTGTCTGAGCTTAACCAAGGCTGAGAGTGTAT 530
Db
344 ArgGlyGlnArgAlaGlyArgProGly-----ArgProArgArgAlaLeuProPro 361
Qy
531 GGACGAGGAGGAGCCCATGAGGAGTGCAGCGTAGACACAGAGGACAGCGCTATG 590
Db
362 GlyAlaGlyAlaGlyGlnProGlu-----Gly 371
Qy
591 AGGAGAGCTGAGAGAGGCGCCAGCAGATGATGATGATGATGATGATGATGATGATGAT 635
Db
372 ArgArgGlyAlaGlySerAlaSerPheGlyTrpArgGlyAspSerLeuArgProGlyGly 391
Qy
636 TGTGCTCAGGATCAAGCGCCCTTGCCTCCAGGTAACACAGATTACAGAGAACTCAA 695
Db
392 -LeuArgArg-----ArgProAlaSerAlaHisArg-ThrGlyLeuArgArgAlaG 409
Qy
696 CTGCAAGAGCC-----AACAGAAATATAGCCAGTGGGCACTTGAAGGGAGATG 746
Db
409 GlyGlnProArgAspArgArgProLeuAla----- 420
Qy
747 GCAGCAGTGGCTGATGAACACATCAATCCAGAAAGCTCAATCTTTCAGTGAAGATT 806
Db
421 -----AlaHisLeuProArgGlnHisLeuArgAlaGlyArgLeuAla 434
Qy
807 TGATTACAGTGGCCA-----TGTCACCCGCTTACACAAAGGAGATGAGG 854
Db
434 rGATSerGlyArgProGlnProAlaLeuCysAspAlaAlaAspGlyLeuArgThrArg 454
Qy
855 CTATGGCGCCCAAGAGAGAACCAAACTGTGAAAGGGCCCAAGCGCTGAGAGCA 914
Db
454 laGlyAlaAspProArgArgGlnProGln-----ProGlyGlyLeuArgSer- 469
Qy
915 CATCTA---CAGGGAATGAT-----GGACATGTGTTTCAATATCTGCAC 956
Db
470 HisAlaGlyAspGlyGluAspProGlyLeuProAlaGlyGlyValAlaThrAlaProArg 489
Qy
957 AATGGCTCGCCACAGAGAGAGA-----TGCCAAGAT-----CCAGGT 992
Db
490 AsnGlyTrpProMetAlaSerGlyArgLeuTrpAsnGlyProGlnGlyProValProGly 509
Qy
993 TACTTTTGGAGATCT-----CTTTGACAGATAGTTCGATTTCAGATAAAGTAGTGG 1046
Db
510 TyrArgTrpSerSerAspArgLeuCysProAspArgSerGly-GlyGluArgGlySileG 529
Qy
1047 CATCTCATGCTG---GCCAGGAACATGACTG----- 1077

Db
529 yProGluGlnArgValThrArgProHisGlyLeuSerAlaGlySerGlyArgTyrAl 549
Qy
1078 -GTAGACTTTGAAGAGAGATGCTA-----TGCAAGCCCGAGATGACCATGTTGT 1127
Db
549 aValGluThrCysGlySileValGlyArgTyrTrpArgSerProArgGluHisAla 569
Qy
1128 GATT 1131
Db
569 eLeu 570
RESULT 11
US-09-252-991A-27062
; Sequence 27062, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27062
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27062
Alignment Scores:
Pred. No.: 0.0081 Length: 741
Score: 125.00 Matches: 83
Percent Similarity: 36.14% Conservatives: 50
Best Local Similarity: 22.55% Mismatches: 166
Query Match: 6.07% Indels: 69
DB: 4 Gaps: 13
US-10-644-659A-1 (1-1146) x US-09-252-991A-27062 (1-741)
Qy 13 GAAAGAAAGAGGGAGGCGCCAGCAAGAGCGCCCTCCGG-----AAGATAGCGACA 66
Db 274 AspArgGlnGlnCysGlnGlyGlnValGluSerThrValGluProAspArgArgThr 293
Qy 67 GCCACCTGCTCATCAGCTTGGCCGAGTGGCAGTGGCGAATGAGAACAGCATC 126
Db 294 AlaGluCysValGlyArgLeuArgGlnArgGlnProGlyGlnHisGlnGlnAlaThr 313
Qy 127 AGGAGGCGCCAGAGCGCTACAGGCTGCTGCCGGAGGAGCCAGGACTCAGCTCAGCT 186
Db 314 GluGlnAlaHis-----GlnGlnValAlaGlnArg 323
Qy 187 CCTAAACCAATCACACCCCTACTTCCACACCAAGAGTCCAGAGTGCCTCCAAAGTCGCCA 246
Db 324 GlnAlaProArgHisProAlaThrGlyHisProGlnArgArgGlnGlyAlaAlaGluVal 343
Qy 247 CCCCCTGCTCCAGAGAGCATGGAGATGGACAAAGCTCAGAGAAAGCCCTCAGGTTTCT 306
Db 344 GlyThrGluHisGlnGlnGlyAspGlyGlnArgAspGlnSerGlyLeuGlyGlnGly 363
Qy 307 CACATCAAAAGAAAGAGGTGTCCAAACGGTGGTCCAGCAAGACTTACGAGAGAGA--- 363
Db 364 ArgHisGlnGlnTrpHisArgAspThrGlyMetAlaGlyProGlyGlnGlnArgGlyGlu 383
Qy 364 -----GGGAC-----GTAGCCACTCAGCCACAGGATGACGAG 396
Db 384 GlnGlnGlyAspGlnArgIleAlaPheGluAlaValAspHisArgAsp-LysGlyTrp 403
Qy 397 AGGATCTCTGCTGCTGTGAACCTGGCGCAGCAGAGATG----- 436

Query Match:	6.05%	Indels:	103
DB:	4	Gaps:	17
US-10-644-659A-1 (1-1146) x US-09-252-991A-25404 (1-256)			
QY	20	AAAGCGGGAGGGCCAGCAGCCCTCCGGAAGATACGACAGCCACCCCTGTGCA	79
DB	1	ArgSerGlyArgThrProThrArgAlaSerGlyGlyArgAsnArgArgAlaArgSer	20
QY	80	TCA-----GTTGGCCCGAGGTGGCAGCAGTGGCGAATGAGAACAGCATCA	127
DB	21	GlyMetProCysArgArgTrpProAlaThrProValAlaArgHisTrpArg	37
QY	128	GGCAGGCCAGCAGCCTACAGCTGGCTGGCGGAGGGACCC-----AGACT	175
DB	38	---ArgLeuArgAlaArgCysGlyThrGlyValProAlaGlyTrpArgThr	56
QY	176	CACCTCAAGCTCTTAACCAATCACACCCCTACTTCACACCAGAAAGCTCAGA	232
DB	57	ArgArgTrpProValaArgAlaAlaProProArgValArgArgGlyPheArgAlaVal	76
QY	233	CCCAAGCTCGCCACCCCGCTGCCAGAAAGCATGGAGTGGACAAAGCTCAGAGAA	292
DB	77	ProProAlaArgHisProGlyAla---AsnGlyCysArgProGlyAlaAlaLeuArg	95
QY	293	CCCTGAGGTTCTCATCATCAAAAGAGAGGTGTCCAAACCGTGTGACGACAGCTT	352
DB	96	ProValArg-----GlyAlaProGlyArgTrpProCysArgGly	108
QY	353	ACGAGAGAGGAGGGAGCGTGGACCCAGCTCAGCAGGAGGAGGATGCTGGTGTGC	412
DB	109	-----AlaGlyArgGlyArgProGlyAlaProValAla	119
QY	413	TTGAACCTGGGAGCCAGAGATGACATTCAGATCTCCAGCCAGCCGCTCCCA	472
DB	120	GlyThrAlaGlyArgArgAlaThr-----AlaSerArgAlaArgProProPro	136
QY	473	CGCGAGGAGAAATGTGCCACCTGTCTGAGTAACTCAAGGCTGGAGAGTGATGG	532
DB	137	ArgAlaArgGlyArgArgCysCysAlaAla---GlyArgAlaGlyAla	153
QY	533	AGCAGGAGGAGCCCATGGAGAGTGTACAGCGTAGACAGAGCAGCGGCTATGAG	592
DB	154	-----GlyProProSerAlaAlaSerArg	161
QY	593	GAGAGCTGAGAGAGGCGCCGAGCAGGATGGAGTGCAGGTGGCTGTGTCAGATCA	652
DB	162	ArgValLeuArgGluSerProGlyArgArgAlaCysArg	174
QY	653	GCCCTTGGCTCCCGAGGTAAACAGATTTACAGAGAAACTCACTGCAAGCCCA	712
DB	175	-----ProProArg-----ValArgSerProProProProArgAsnArg	187
QY	713	ATATAGCCAGTGGCACTTGAAGGAGAGTGCAGCAGTGGCTGATGACACATAC	772
DB	188	ArgCysAlaGlyTrpSerProThrAlaGlyAsp-----	198
QY	773	AATCCCAAGAGCTCAATCTTTTTCAGTGAAGAGTTTATTACGAGCTGGCCAT	832
DB	199	ArgProGlyAlaSer-----ArgProPro	206
QY	833	GCTACACAAAGAGAGTGGGCTATGGCGCCCAAGAGGAGAACCAAACTGTGAA	892
DB	207	-----ProArgAlaProProProAlaProProArgSerArgArg	219
QY	893	GGGCAAGCGTGTGTGAGGAGCATCTACAGGAAATGATGGACATGTCTTATTCT	952
DB	220	Gly-----ThrSerSerArgTrpThr-----	226
QY	953	GCAATAGG 961	
DB	227	AlaGlnTrp 229	

Db	403	GGlyLeuAspLeuGlnGlyMetAlaGluGlnArgGlnArgGlnGlnHisGlnAlaG	423
QY	437	-----ACATTGACAGAACTCTCCACAGC	459
Db	423	userAspGlnAlaLeuAlaLysAlaLeuGlnAlaArgThrValAlaGlnAlaGlnG	443
QY	460	CACGGTCCCAACCGGAGGAGAAATGTGCCAAGCTGTGTCTGAGCTAACCAAGGC	519
Db	443	nGlnAlaGlyGlnGlnGlyGlnArgGlnGlnArgGlyGlnGlnGlnGlnGlnG	463
QY	520	TGGAGAGTGTGAGCAGGAGGAGCCACATGGAGGAGTGCACAGCTACACAGAGG	579
Db	463	ishisGlnSerGlyAlaAspValGlyAlaGlnHisGlnArgGlnAlaG	483
QY	580	AGCGGCTATGGAGAGGCTGGAGAGGCGCCGAGCAGGATGGAGTGGCTGTG	639
Db	483	InglyAlaThrGly-----GlyGlnGlyArgCysHisGlnThrGlyGlyAla	500
QY	640	GTCAGGATCAAGGCCCTTGCCTCCAGGTAACAGATTTACAGAGAACTCAACTGC	699
Db	500	laLeuGlnGlnAlaGlyAspAlaGlnProGlyGlnGln-----GlyGlnProAlaVala	519
QY	700	AAAGCCCAACAGAAATATAGCCAGTGGCACTTGAAGGAGAGTGGCAGCAGTGC	756
Db	519	InValAlaAlaGlu-----ProValAlaGlnArgSerAlaGlnThrAla-GlnHisAla	536
QY	757	GCTGATGAACATACATACCTCCCTCCAGTCAATCTTTCAGTGAAGAGTGTGATTAC	813
Db	537	AlaGlyAspHisValArgAlaProGlnGlnGlyHisGlyThrGlyGlnValaspGln	556
QY	814	GAGTGGCCATGTCCACCGCTTACACAAAGGAGATGAGGC-----	855
Db	557	Glu-----AspGlySerGlnHisGlyThrProGlySerGluTrpIleValLys	573
QY	856	-----TATGGCGCGCCCAAGAGAACTGCTGAAAGGCGC	897
Db	574	GlyThrGlySerLeuThrHisGlyGlnProGluArgGly-----ThrValGluArgAla	591
QY	898	AAGCGTGTGAGGACATCTACAGGAAATGATGACATGTGCTTCATTATCTGACAC	957
Db	592	ArgAlaValGlnGlnValPheHisPheAspLeuAlaGlnAlaAlaGlyValCysAla	611
QY	958	ATGGCTCCACACAGCA 975	
Db	612	HisAlaAspHisHisGln 617	
RESULT 12			
US-09-252-991A-25404			
Sequence 25404, Application US/09252991A			
Patent No. 6551795			
GENERAL INFORMATION:			
APPLICANT: Marc J. Rubenfield et al.			
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS			
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS			
FILE REFERENCE: 107196.136			
CURRENT APPLICATION NUMBER: US/09/252,991A			
CURRENT FILING DATE: 1999-02-18			
PRIOR APPLICATION NUMBER: US 60/074,788			
PRIOR FILING DATE: 1998-02-18			
PRIOR APPLICATION NUMBER: US 60/094,190			
PRIOR FILING DATE: 1998-07-27			
NUMBER OF SEQ ID NOS: 33142			
SEQ ID NO 25404			
LENGTH: 256			
TYPE: PRT			
ORGANISM: Pseudomonas aeruginosa			
US-09-252-991A-25404			
Alignment Scores:			
Pred. No.:	0.000545	Length:	256
Score:	124.50	Matches:	78
Percent Similarity:	31.27%	Conservative:	23
Best Local Similarity:	24.15%	Mismatches:	119

RESULT 13

US-09-252-991A-17580
 ; Sequence 17580, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR FILING DATE: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 17580
 ; LENGTH: 482
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-17580

Alignment Scores:

Pred. No.: 0.000739 Length: 482
 Score: 124.50 Matches: 107
 Percent Similarity: 35.8% Conservative: 40
 Best Local Similarity: 26.10% Mismatches: 150
 Query Match: 6.05% Indels: 113
 DB: 4 Gaps: 22

US-10-644-659A-1 (1-1146) x US-09-252-991A-17580 (1-482)

QY 14 AAAAGGAAGCGGAGGCGCCAGCAGAGCGCCCTCCGGAAGATAGCAGCAGCCACCC 73
 DB 86 ArgArgSerAlaGlnProAlaGlyProArgGlnAlaAlaAspArgGluHisProGly 105
 QY 74 TGGTCATCAGCTTGGCCCGAGCTGGCGAGCAGTGGCGGAATGAGACAGCATCAGCAGG 133
 DB 106 AsnAlaAlaValArgProGluProGlyAlaGluGlyAspGlnGlyProAlaArgSerHis 125
 QY 134 CCCAGGAGCCCTACAGGCTGGTGGTCCGG----- 160
 DB 126 ---ArgAlaLeuArgAlaGlyAlaArgLeuArgGlnProArgProAlaGluGlyHisPro 144
 QY 161 ---GAGGAGCCAGGAGTCTACCTCAAGCTCTAAACCAATCACACCCCTACTTCAACC 217
 DB 145 GlyValGlyArgAlaHisArgLeuProGlyAspProAlaGlyProAspProGluSer 164
 QY 218 AGAAGCTCAGATGCCCAAGTCGCCACCCCGCTGCCAGAGGACATGAGATGGAC 277
 DB 165 ArgProGlyGluLeuProAlaValAlaHisAlaArgArgArgLeuileGlyAsnHisAsp 184
 QY 278 AAAGCT-----CAGAGAAGCC----- 295
 DB 185 GluAlaProAlaAlaCysProValGlyAlaPhePheLeuProAlaArgGlyProGlnAla 204
 QY 296 CTGAGTTCTC----- 307
 DB 205 LeuArgGlyLeuArgArgThrArgSerAlaArgAlaSerAsnArgGlnArgProThr 224
 QY 308 ---ACATCAAAAGAGAGAGGTGCCAAACCGGTGGTCAGCAAGA----- 349
 DB 225 SerThrSerAsnAlaGlnArgArgProIleSerTrpAlaSerArgProIleSerGlyGly 244
 QY 350 -----CTTACAGAGAGAGGGGACGTGAGCCACCTCAGCCACA 388
 DB 245 ProIleSerSerProProTyProThrValGlu-----ThrAlaAlaThrProIleThr 262
 QY 389 GGTACGAGA---GGATGCTG-----GTGTGCTTGAACCTGGCCAGCAGAGAAATGACA 439
 DB 263 GlySerMetProGlyTrpLeuProAlaAlaLeuIleThrIleGly-----Thr 278
 QY 440 TTGACAGAAATCTCCACAGCCACGCGTCCCAACAGCGGA-----GGA 481

DB 279 IleThrAspSerProArgProThrThrProLysProAlaMetLeuSerAlaGlyGlnGly 298
 QY 482 GAAATGTGCCAACCTGGTGTCTGAGC-----TAACCAAGGCTCGAGAG 526
 DB 299 LeuAsnSerAlaThrSerMetProSerValAlaSerGluProArgProArgSerAsnGly 318
 QY 527 TGATGGAGC-----AGGAGAGCCCATCGAGGAGTGGAGGAGTACAGCTAGACAG 574
 DB 319 AlaSerSerMetArgProArgArgSerProAsnArgProThrThrIleAlaAla 338
 QY 575 AGGACACGCGCTATGGAGAGAGGCTGAGAGAGG-----CCGAGCAGGATGGAGTGCAGG 631
 DB 339 GluAsnAlaAlaLysProSerAlaAlaThrArgSerGluAlaSerArgThrLeuCysArg 358
 QY 632 TGGCTGTGTCAGGATCAGCGCCCTTGCCTCCAGGTAACAGATTACAGAGAAC 691
 DB 359 -----AsnSerAlaLeuGlnSerAlaThrProProSerGluThrAsnAlaGlnArgIle 376
 QY 692 TCAACTGCAAGGCCCAACAGAAATATAGCCAGTGGGCAACTTGAAG----- 739
 DB 377 SerArgProArgProSerLys-----AlaArgGlyLysValLysProProLeu 394
 QY 740 -----GGAGATGGCAGCAGTGGCTGATGAACATACATATCCACA 781
 DB 395 ArgAlaProAlaLeuLeuGlySerGlyGlySerAsn---AlaArgThrLysAsnGlyArg 413
 QY 782 AGCTCAATCCTTTCAAGTGAAGAGTTGATTACAGAGTGGCCATGT-----CCA 829
 DB 414 -----LeuSerArgProSerAlaValArgMetArgLysCysGlnAsnGlyValSer 430
 QY 830 CCGCCTTACACAAGAGAGATGAGGCTATGGCGCCCAAGAGAGAACCAAACTG--- 886
 DB 431 ProAla-----ArgAlaAlaAlaLeuAsnAlaProThrArgGlnProArgLeuTy 447
 QY 887 -----CTGAAAGGCCCAAGCGTCTGAGGAGCAGCATCTACAGGAAA 928
 DB 448 MetProCysSerMetAspMetGlnGlyLeuProSerArgArgSerThrThrAlaTrp 467
 QY 929 TGATGGACATGCTTCATTATCTGCACA 958
 DB 468 MetPheIleAlaThrSerIleProAlaLys 477

RESULT 14

US-09-252-991A-29215
 ; Sequence 29215, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 29215
 ; LENGTH: 1113
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-29215

Alignment Scores:
 Pred. No.: 0.00124 Length: 1113
 Score: 124.00 Matches: 92
 Percent Similarity: 34.41% Conservative: 25
 Best Local Similarity: 27.06% Mismatches: 89
 Query Match: 6.03% Indels: 134
 DB: 4 Gaps: 18

US-10-644-659A-1 (1-1146) x US-09-252-991A-29215 (1-11113)

QY 6 TCGGGCGAAGAAAGCGG-----GGAGGCGCCAGCAAGCGCCCTCGC 53
Db 81 AlalGlyArgThrGlyLeuArgHisGlnProHisGlyGlyAsnAlaAlaGluArgProGly 100
QY 54 GAAGATACGCACACGCCAC---CCTGGTCATCAGCTTGGCCCGAGGTGGCAGCGTGGC 110
Db 101 GlyAsnAlaAlaGlnHisArgProGlyHisArg-----GlyArgThrAlaGluProGly 118
QY 111 GAATGAGAA-----CAGCATCAGCGAGCGCCAGGCTTACAGGTGGCTGCGCGGAGG 164
Db 119 AsnHisArgAspProGluHisArgAlaGlyProGlyAla-GlnGlyGly---ProGlyG 137
QY 165 GACCCAGGATTCACCTCAAGCTCTAAACCAATCACACCCCTTACTTACACAGAAAGC 224
Db 137 yGluGlnAspGlnValArgProArgGlnHisGluProArgAspProHisProAlaG 157
QY 225 TCAGAGTGCCTCAAGTCGCCACCCCGCTGCCAGAGGACATGGAGATGCACAAAGCTC 284
Db 157 n-ArgHisProArgPheArgPheLeuGluAspArgGlyArgGlnAlaGlySerGlyLys 211
QY 372 -----GAGCCACCTCAGCCA-----CAG 389
Db 212 ProProPheGlnSerProArgProAspProGlyArgProAspHisAlaGlySerGlyArg 231
QY 390 GTACGAGGAGGATGCTGTGTGCTTGAACCTGGCGAGCCA----- 429
Db 232 ProArgGluAla-ThrGlyThrGlyGlnProGlyLeuProGlyTyrProAspProIleAl 251
QY 430 -----GAGATGACATTGAC----- 444
Db 251 aGlyArgProAlaAlaAlaGluAlaAspProHisGlnProGlyArgGlnArgHisGlnVa 271
QY 445 -----AGATCTCTCAGCCAGCCAGCGCTCCCAACCGCGAGGAG 482
Db 271 LHisProGlyArgHisArgArgAlaArgHisAlaArgGlyArgLysArgArgProAr 291
QY 483 AAAATGT-----GCCAA 494
Db 291 GaAlaAlaAspGlnArgThrGlyHisArgTyrArgProLeuArgGlyGlyProAlaSe 311
QY 495 CCTGGTGTCTGAGCTAAACCAAGGCG-----TGG----- 522
Db 311 rLeuValGlnGlyLeuGlnProGlyArgGlnLeuThrValAlaThrSerArgTyrHisAr 331
QY 522 ----- 522
Db 331 gProGlyProGlyAspLeuGlnAlaProAspArgAlaAspGlyArgArgAspArgArgAr 351
QY 523 -----AGAGTGATGGAGCAGGAGGAGCCACATCGAGGAGTGA 560
Db 351 gGlnTyrAlaArgGlyArgArgValLeuAspGlnProGluSerAlaGluLysSer-- 370
QY 561 CAGCGTAGACACAGAGGACAGCGGCTATGGAGGAGGCTGAGGAGAGCGCCGACGAGA 620
Db 371 -----ArgArgGlnArgGlyAlaGly-ArgLeuLeuGlyArgGlyProThrArg- 386
QY 621 TGGAGTGCAGGTGGCTGTGTGTCAGGATCAAGCGCCCTTGCCTCCCA 668
Db 387 GlyAlaAlaArgThrAlaGlyThrAsp---AlaProLeuAlaAlaPro 401

RESULT 15

US-09-252-991A-26942

; Sequence 26942, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196-136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 26942

; LENGTH: 437

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-26942

Alignment Scores:

Pred. No.: 0,00142 Length: 437

Score: 121.50 Matches: 48

Percent Similarity: 37.85% Conservative: 19

Best Local Similarity: 27.12% Mismatches: 56

Query Match: 5.90% Indels: 54

DB: 4 Gaps: 8

US-10-644-659A-1 (1-1146) x US-09-252-991A-26942 (1-437)

QY 3 GGCTCGGCGAAGAAAGCGGAGGCGCCAGGAGCGCCCTCCGAGAT--- 59

Db 250 GlyGlnGlyArgLeuGlyArgThrValAlaAspGlnProGlyAspProGlyAspHis 269

QY 60 -----ACGCACAGCCACCCCTGCT----- 77

Db 270 GlyThrAspLeuAlaHisProHisProGlyLeuArgArgSerArgArgHisArgArg 289

QY 78 ---CATCAGCTGGCCCGAGGTGGCAGCAGTGGCGGATGAGACACATCAGCAGGC 134

Db 290 ProHisArgSerGlyArgThrValAlaAla-----ProProArgArgThrArg 306

QY 135 CCAGGAGCCTACAGG-----CTGGCTCGCGGAGGAGCCAGGACTCACCTCAAGCTCC 188

Db 307 ProGlyProHisArgAspProLeuAlaAlaGlyArgArgProGly----- 321

QY 189 TAAACCAATCACACCCCTTACTTACACCAAGAGCTCAGTGCCTCCAAAGTCCGCC 248

Db 321 ----- 321

QY 249 CCGCTCGCCAGGAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCTGAGGTTCTCA 308

Db 322 ProProAlaProValArgAlaGlnGlyThrArgGlnGlnArgAsnProTyrGlyArg 341

QY 309 CATCAAAAGAAAGAGGTGTCCAAAACGGTGTTCAGAAAGACTTACGAGAGAGGAGGA 368

Db 342 HisProGlnGlnGlyProAlaAspArgGlyArgProHisAlaArgArgArg--- 360

QY 369 CGTGAGCCACCTCAG-----CCACAGGTACGAGCGGATGCTGTGTCTTGAACCTGG 422

Db 361 -----ProAspGlnGlyProGlyProArgGlu-GluThrGly-----GlnProG 376

QY 423 GCAGCCAGAGAATGCATTCACAGAATCTCCACAGCCACGGCTCCCA 471

Db 376 YArgProGluLeuAsnValProSerMetGluAsnGluAsnGlyAspPro 392

Search completed: May 4, 2004, 08:56:46

Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2004, 00:51:08 ; Search time 525 Seconds
(without alignments)
9874.076 Million cell updates/sec

Title: US-10-644-659A-1
Perfect score: 1146
Sequence: 1 atggctccggcgaaagga.....tgattacgctactcaagtga 1146

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2936184 seqs, 2261732022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications NA:*
- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:*
 - 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
 - 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
 - 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
 - 6: /cgn2_6/prodata/2/pubpna/PCRU_PUBCOMB.seq:*
 - 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
 - 8: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:*
 - 9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:*
 - 10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:*
 - 11: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
 - 12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
 - 13: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:*
 - 14: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:*
 - 15: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
 - 16: /cgn2_6/prodata/2/pubpna/US10D_PUBCOMB.seq:*
 - 17: /cgn2_6/prodata/2/pubpna/US10E_NEW_PUB.seq:*
 - 18: /cgn2_6/prodata/2/pubpna/US10F_NEW_PUB.seq:*
 - 19: /cgn2_6/prodata/2/pubpna/US10G_PUBCOMB.seq:*
 - 20: /cgn2_6/prodata/2/pubpna/US10H_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1144.4	99.9	1322	16	US-10-104-047-1199
2	499.4	43.6	537	9	US-09-864-761-8525
3	363.4	31.7	365	9	US-09-864-761-25251
4	198.2	17.3	388	16	US-10-191-803-820
5	42.8	3.7	594	13	US-10-142-426-10
6	42.8	3.7	594	15	US-10-123-155-10
7	42.8	3.7	594	15	US-10-146-731-10
8	42.8	3.7	594	15	US-10-140-472-10
9	42.8	3.7	594	15	US-10-141-761-10
10	42.8	3.7	594	15	US-10-142-885-10
11	42.8	3.7	594	15	US-10-158-790-10
12	42.8	3.7	594	16	US-10-137-871-10
13	42.8	3.7	594	16	US-10-140-923-10
14	42.8	3.7	594	16	US-10-141-756-10

15	42.8	3.7	594	16	US-10-141-759-10	Sequence 10, Appl
16	42.8	3.7	594	16	US-10-140-805-10	Sequence 10, Appl
17	42.8	3.7	594	16	US-10-140-864-10	Sequence 10, Appl
18	42.6	3.7	671	15	US-10-184-644-346	Sequence 346, App
19	42.6	3.7	671	15	US-10-184-634-346	Sequence 346, App
20	40.4	3.5	14800	9	US-09-954-456-1601	Sequence 1601, App
21	40.4	3.5	14800	15	US-10-269-909-61	Sequence 61, Appl
22	39.2	3.4	1046	13	US-10-425-114-23019	Sequence 23019, A
23	38.8	3.4	14835	15	US-10-240-965-113	Sequence 113, App
24	38.6	3.4	5731	17	US-10-641-643-1415	Sequence 1415, App
25	38.6	3.4	13210	10	US-09-814-353-22015	Sequence 22015, A
26	38.6	3.4	14255	9	US-09-964-824A-282	Sequence 282, App
27	38	3.3	732	13	US-10-424-599-29937	Sequence 29937, A
28	37.8	3.3	711	15	US-10-029-386-20329	Sequence 20329, A
29	37.8	3.3	1140	13	US-10-424-599-42070	Sequence 42070, A
30	37.4	3.3	486	9	US-09-938-842A-355	Sequence 355, App
31	37.4	3.3	486	11	US-09-938-842A-355	Sequence 355, App
32	37.4	3.3	492	10	US-09-770-961-618	Sequence 618, App
33	37.2	3.2	954	15	US-10-156-761-3934	Sequence 3934, App
34	37.2	3.2	235070	13	US-10-087-192-1990	Sequence 1990, App
35	37.2	3.2	9025608	15	US-10-156-761-1	Sequence 1, Appl
36	37	3.2	471	10	US-09-918-995-29236	Sequence 29236, A
37	37	3.2	1642	13	US-10-302-172-873	Sequence 873, App
38	36.4	3.2	999	15	US-10-184-644-434	Sequence 434, App
39	36.4	3.2	999	15	US-10-184-634-434	Sequence 434, App
40	36.4	3.2	1180	16	US-10-260-238-754	Sequence 754, App
41	36.2	3.2	660	13	US-10-142-426-480	Sequence 480, App
42	36.2	3.2	660	15	US-10-123-155-480	Sequence 480, App
43	36.2	3.2	660	15	US-10-146-731-480	Sequence 480, App
44	36.2	3.2	660	15	US-10-140-472-480	Sequence 480, App
45	36.2	3.2	660	15	US-10-141-761-480	Sequence 480, App

ALIGNMENTS

RESULT 1

US-10-104-047-1199 ; Sequence 1199, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1199
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-104-047-1199

Query Match 99.9%; Score 1144.4; DB 16; Length 1322;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1145; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	ATGGCTCCGGCGAAAGGAAAGCGGGGCGCCAGCCAGAGCGCCCTCCGGAAGATA 60
DB	57	ATGGCTCCGGCGAAAGGAAAGTGGGAGGCGCCAGCCAGAGCGCCCTCCGGAAGATA 116
QY	61	CGCACAGCCACCTGGTGTATCATAGCTTGGCCCGAGGTTGGCAGCAGTGGCGGAATGAGAAC 120
DB	117	CGCACAGCCACCTGGTGTATCATAGCTTGGCCCGAGGTTGGCAGCAGTGGCGGAATGAGAAC 176
QY	121	AGCATCAGGAGCGCCAGAGGCTACAGCTGGTGGCGGGAGGAGCCAGAGACTCACCT 180
DB	177	AGCATCAGGAGCGCCAGAGGCTACAGCTGGTGGCGGGAGGAGCCAGAGACTCACCT 236
QY	181	CAAGCTCTTAACCAATCATACACCCCTACTTACACACAGAAAGCTCAGAGTGGCCCCAAG 240

Db	237	CAAGCTCCTAAACAAATCACACCCCTTCTTACACAGAAAGCTCAGAGTGCCTCCCAAG	296
Qy	241	TCGCCACCCCGCTGCCAGAGGACATGGAGATGGCAAAAGCTCAGAGAAAGCCCTGAG	300
Db	297	TCGCCACCCCGCTGCCAGAGGACATGGAGATGGCAAAAGCTCAGAGAAAGCCCTGAG	356
Qy	301	GTCTTCTACATCAAAAGAAAGAGTGTCCAAACCGTGTGACGAGACTTACGAGAGA	360
Db	357	GTCTTCTACATCAAAAGAAAGAGTGTCCAAACCGTGTGACGAGACTTACGAGAGA	416
Qy	361	GGAGGGACGTGACCCACCTCAGCCACAGGTATCAGAGGGGATGCTGTGTGCTGAACCT	420
Db	417	GGAGGGACGTGACCCACCTCAGCCACAGGTATCAGAGGGGATGCTGTGTGCTGAACCT	476
Qy	421	GGGAGCCAGAGATGATTCAGAGATCTCTCCACAGCCACCGCTCCCAACGGGAGG	480
Db	477	GGGAGCCAGAGATGATTCAGAGATCTCTCCACAGCCACCGCTCCCAACGGGAGG	536
Qy	481	AGAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCAGG	540
Db	537	AGAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCAGG	596
Qy	541	GAGCCACATGGAGAGTGAAGCGTATGACAGAGACAGCGGCTATGGAGAGGCT	600
Db	597	GAGCCACATGGAGAGTGAAGCGTATGACAGAGACAGCGGCTATGGAGAGGCT	656
Qy	601	GAGGAGGCGCGGAGCAGATGGAGTGCAGGTGCTGTGTGCTGAGATCAAGCGCCCTTG	660
Db	657	GAGGAGGCGCGGAGCAGATGGAGTGCAGGTGCTGTGTGCTGAGATCAAGCGCCCTTG	716
Qy	661	CCCTCCAGGTAAACAGATTTACAGAGAACTCACTGCAAGGCGGCAACAGAAATATAGC	720
Db	717	CCCTCCAGGTAAACAGATTTACAGAGAACTCACTGCAAGGCGGCAACAGAAATATAGC	776
Qy	721	CCAGTGGCAACTTGAAGGGAGATGGCAGAGTGGGCTGATGAACACATCAATCCAG	780
Db	777	CCAGTGGCAACTTGAAGGGAGATGGCAGAGTGGGCTGATGAACACATCAATCCAG	836
Qy	781	AACTCAATCTCTTTCAGTGAAGAGTTTGAATACGAGTGGCCATGTCACCCGCTACAC	840
Db	837	AACTCAATCTCTTTCAGTGAAGAGTTTGAATACGAGTGGCCATGTCACCCGCTACAC	896
Qy	841	AAAGAGATGAGGCTGTGGCGCCCAAGAGGAACCAAACTGCTGAAGGGCCCAAG	900
Db	897	AAAGAGATGAGGCTGTGGCGCCCAAGAGGAACCAAACTGCTGAAGGGCCCAAG	956
Qy	901	CGTGTGTGAGGACACATCTACAGGGAATGATGGACATGTGCTTCATTTATCTGCACATG	960
Db	957	CGTGTGTGAGGACACATCTACAGGGAATGATGGACATGTGCTTCATTTATCTGCACATG	1016
Qy	961	GCTCGCCACAGAGATGGCAAGATCCAGGTACTTTTGGAGATCTCTTTGACAGATAC	1020
Db	1017	GCTCGCCACAGAGATGGCAAGATCCAGGTACTTTTGGAGATCTCTTTGACAGATAC	1076
Qy	1021	GTTCGTATTTTCAGATAAAGTGTGGGCACTTCTCATGCGTGCAGGAAACATGATGTA	1080
Db	1077	GTTCGTATTTTCAGATAAAGTGTGGGCACTTCTCATGCGTGCAGGAAACATGATGTA	1136
Qy	1081	GACTTTGAGGAGAGATGCTATGGCAAGCCGAGATGACCAATGTTGTGATTCGCTACTC	1140
Db	1137	GACTTTGAGGAGAGATGCTATGGCAAGCCGAGATGACCAATGTTGTGATTCGCTACTC	1196
Qy	1141	AACTGA 1146	
Db	1197	AACTGA 1202	

RESULT 2
US-09-864-761-8525/c
; Sequence 8525, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeonica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 8525
LENGTH: 537
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC023344.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
US-09-864-761-8525
Query Match 43.68; Score 499.4; DB 9; Length 537;
Best Local Similarity 99.8%; Pred. No. 6.1e-152;
Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 169 CAGGACTCACCTCAAGCTCCTAAACCAATCACACCCCTACTTTCACACCAAGAGCTCAG 228
Db 537 CAGGACTCACCTCAAGCTCCTAAACCAATCACACCCCTACTTTCACACCAAGAGCTCAG 478
Qy 229 AGTGCCCAAGTCCGACCCCGCTCCAGAGGACATGGAGTGCACCAAGCTCAGAG 288
Db 477 AGTGCCCAAGTCCGACCCCGCTCCAGAGGACATGGAGTGCACCAAGCTCAGAG 418

391	QY	TACGAGAGGGATGCTGCTGCTTGCTTGAACCTGGGCACCCAGAGAAATGACATTGACAGATC	450
226	Db	TTTGAGDGTFTMLRDKDKEBAEAIKHALEBEKAMTSGRRSRQRQRRFEKRLRKLIS	285
451	QY	CTCCACAGCCAGCGCTCCCCAACGGGAGGAGAAAATGTGCCAACCTGTGTCTGAGCTA	510
286	Db	PPSYARRDSPTYDKTSPSESSSESRSRSTPQREEKITFTTSGDSEAAAAAAA	345
511	QY	ACCAAGGCGCTGGAGAGTGTATGGAGCAGGAGGAGCCACATGGAGGAGTGCACAGCGTAGAC	570
346	Db	AAAGVTGTKPPAPPPQGGPAPGENASARRSSSSSSSSASRTSSRSSSSRSRRG	405
571	QY	ACAGAGACAGCGCTATGGAGGAGGCGTGCAGAGAGGCCCGCAGCAGATGGAGTGC---	628
406	Db	GGYRSGSHAKRSRSRSRSRSTRYSRSRGRHSGGSGRDGHRYSRSPARRGGYGP	465
629	QY	-AGTGGGTGTGTCAGATCAACAGCCCGCTTCCTCCACAGTAAACAGATTTACAGAG	687
466	Db	RRRSRSRSHSGDVRRRGGRLRHSSRSRSSSWSLSPRSRLTRSRSHSPSPSQSRRS	525
688	QY	AAACTCAACTGCAAGCCCAACAGAAATATAGCCAGTGGCAACTTGAAGAGGAGATGG	747
526	Db	RRSQSPSPSPAREKLTFRPLAASFVAGEKLLKKTTFPAAGKETGAAKVTQADPSGEATEDEAE	585
748	QY	CAGCAGTGG	756
586	Db	GAEQAVQGG	594

```

RESULT 8
US-10-140-472-10
; Sequence 10, Application US/10140472
; Publication NO. US20030138886A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C168
; CURRENT APPLICATION NUMBER: US/10/140,472
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-472-10

```

```

Query Match      3.7%; Score 42.8; DB 15; Length 594;
Best Local Similarity 10.04; Pred.No. 0.0057;
Matches 37; Conservative 152; Mismatches 177; Indels 3; Gaps 1
QY 391 TAGAGAGGGANGCTGTGTGTTGAACCTGGGCAGCCAGAAATGACATTCACAGAAATC 450
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 226 TTYGMAADGDFVFMRLRKDKKEAAEAIKHKALDEEFGMYSGRRSRQREFREKRLGRKIS 285
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
QY 451 CTCACAGCCAGCGGTCTCCCAACCGGAGGAGAGAAATGTGCCAACTCGTGTCTCAGGTA 510

```

286	DB	PSYARRSDPYDPYKRSPPSSSSSSSSSRSSRSPPTGSEKKITITTSFGSGDEAAAAA	345
511	QY	ACCAAGGGCTGGAGAGTGTATGACAGAGAGAGCCCATGTGAGGAGTGCACGCTAGAC	570
346	DB	AAASGVTTGKPPAPPQPGGAPGNARRSSSSSSSSSSSSASRTSSRSSRSSRSRG	405
571	QY	ACAGAGGACAGCGGCTATGGAGGAGAGCTGAGAGAGGCCGAGCAGGATCGAGTGC--	628
406	DB	GGYRSRGHARSRSWRSRSTRYRSRGRHSGGSRDGHYRSRPARRGVGP	465
629	QY	-AGGTGGCTGTGTCAGGATCAAGCGCCCTTGCCCTCCAGAGTAACAGATTTACAGAG	687
466	DB	RRRSRSRSHSGDVRPGCGGLRHHSRNRSSNLSPSRSRLTRNSHSPSPQSRSRS	525
688	QY	AAACTCACTGCAAAGCCCAACAGAAATATAGCCAGTGGCGCACTTGAAGGGAGATGG	747
526	DB	RRSQSPSPSPAREKLTIRPAASPVGKLTKEPAAGKETGAAKVTQADASGEAETDAE	585
748	QY	CAGCAGTGG	756
586	DB	GAEQAVQGG	594

```

RESULT 9
US-10-141-761-10
/ Sequence 10, Application US/10141761
/ Publication NO. US20030148432A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Beresini, Maureen
/ APPLICANT: DeForge, Laura
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Filwaroff, Ellen
/ APPLICANT: Gao, Wei-Qiang
/ APPLICANT: Gerritsen, Mary E.
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Sherwood, Steven
/ APPLICANT: Smith, Victoria
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Watanabe, Colin K
/ APPLICANT: Wood, William
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE OF INVENTION: ACIDS ENCODING THE SAME
/ FILE REFERENCE: P3330R1C198
/ CURRENT APPLICATION NUMBER: US/10/141,761
/ CURRENT FILING DATE: 2002-05-08
/ Prior application removed - See Palm or File Wrapper
/ NUMBER OF SEQ ID NOS: 550
/ SEQ ID NO 10
/ LENGTH: 594
/ TYPE: PRT
/ ORGANISM: Homo Sapien
US-10-141-761-10

```

```

Db      346  AAASGVTTGKFPAPPPQCGPAPGNAGARRRRSSSSSSSSSSASASRTSSSSSSSSSSSSRRRG 405
QY      571  ACAGAGGACAGCGGCTATGGAGGAGAGGCTGAGGAGAGAGCCCGAGCAGGATGGAGTGC-- 628
Db      406  GGYRSRGHARSRRSSRSMRSRRRYRSRSGRRHSRGGSRDGHYRSRSPARRGGYGP 465
QY      629  -AGGTGGCTGGTTCAGGATCAACGCGCCCTTCCTCCCTCCAGGTAACACAGATTTACAGAG 687
Db      466  RRRSRSRSHSGDVRVARGRGLRHHSRRSSRSSLSPFSRSLRTRSRSHSPSPSQRSSS 525
QY      688  AAACTCAACTGCAAGCCCAACAGAAAATATAGCCAGTGGCGCAACTTGAAGGGAGATGG 747
Db      526  RSRSSQSPSPAREKLTRPAASPAVGEKLNKTEPAAGKETGAAKVTQADASGEAETDAE 585
QY      748  CAGCAGTGG 756
Db      586  GAEQAVQGG 594

RESULT 10
US-10-142-885-10
; Sequence 10, Application US/10142885
; Publication No. US20030157604A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumaas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C248
; CURRENT APPLICATION NUMBER: US/10/142,885
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-142-885-10

```

	Query Match	3.7%	Score 42.8;	DB 15;	Length 594;
	Best Local Similarity	10.0%;	Pred. No. 0.0057;		
Matches	Conservative 152;	Mismatches 177;	Indels	Gaps	1;
QY	391 TACGAGGGGATCGTGTGGCTTGTAACCTGGGCACCAGAGAATGACATTGACAGATC	450			
Db	226 ITYGWADGDVFVRLRKDKKEAEAIKHALEEKANYSGRSRRRRFRFRKRLEGRKIS	285			
QY	451 CTCACAGCCAGGGCTCCCACGCCGAGGAGGAANAATGTGCCAACTCGGTGTCTGAGCTA	510			
Db	286 PPSYARRDSPVDYFKRPSSSESSRSRSTPTGREBKITFTIFSGSDEEAAAAAAA	345			
QY	511 ACCAAGGCGCTGGAGAGTGTGTCAGCAGGAGGCCCAATGGAGGAGTAGTGACAGGCTAGAC	570			

[illegible]

Search completed: May 5, 2004, 03:14:38
Job time : 558 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2004, 11:56:32 ; Search time 495 Seconds
(without alignments)
9835.222 Million cell updates/sec

Title: US-10-644-659A-1
Perfect score: 1146
Sequence: 1 atgctccggcgaaagga.....tgattacgtactcaagtga 1146

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1980s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002s.*
- 7: Geneseqn2003as.*
- 8: Geneseqn2003bs.*
- 9: Geneseqn2003cs.*
- 10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1144.4	99.9	1322	9	ADB63045
2	948.4	82.8	1373	5	AAS92182
C 3	499.4	43.6	537	4	AAI17811
C 4	499.4	43.6	537	4	ABA62763
C 5	499.4	43.6	537	4	AAI42779
C 6	499.4	43.6	537	4	ABA30059
C 7	499.4	43.6	537	4	AAK36965
C 8	499.4	43.6	537	4	AAK11153
C 9	499.4	43.6	537	4	ABS36644
C 10	363.4	31.7	365	4	AAI27007
C 11	363.4	31.7	365	4	ABA75275
C 12	363.4	31.7	365	4	AAI5838
C 13	363.4	31.7	365	4	ABA39931
C 14	363.4	31.7	365	4	AAK49906
C 15	363.4	31.7	365	4	AAK23820
C 16	363.4	31.7	365	4	ABS49550
C 17	205	17.9	261	3	AAC00343
C 18	205	17.9	265	2	AAK40341
C 19	56.8	5.0	10732	3	AAI10594
C 20	55	4.8	552	4	ABLO4119
C 21	55	4.8	2610	4	ABLO4118
C 22	52.2	4.6	2000	7	ADA71938
C 23	49.8	4.3	1649	4	ABLO6465

24	49.8	4.3	7090	4	ABLO6464
C 25	42.4	3.7	2896	7	ACC47372
C 26	42.4	3.7	3054	7	ACC47371
C 27	40.4	3.5	14800	6	ABL66291
C 28	39.2	3.4	2000	7	ADA71938
C 29	39	3.4	6741	3	AAI10595
C 30	38.8	3.4	14835	6	AAS94858
C 31	38.8	3.4	28564	9	ADA71107
C 32	38.8	3.4	28564	9	ADA71113
C 33	38.8	3.4	28564	9	ADA71117
C 34	38.8	3.4	28564	9	ADA71111
C 35	38.8	3.4	28564	9	ADBE83334
C 36	38.8	3.4	28564	9	ADE57550
C 37	38.8	3.4	28564	9	ADE57554
C 38	38.6	3.4	6940	6	ABS73156
C 39	38.6	3.4	11910	9	ADC27511
C 40	38.6	3.4	12042	5	AAS76999
C 41	38.6	3.4	14255	2	AAQ43526
C 42	38.6	3.4	14255	2	AAQ75181
C 43	38.6	3.4	14255	2	AAV20477
C 44	38.6	3.4	14255	6	ABL67248
C 45	38.6	3.4	14255	6	ABS73200

ALIGNMENTS

RESULT 1
ADB63045
ID ADB63045 standard; cDNA; 1322 BP.
XX
AC ADB63045;
DT 04-DEC-2003 (first entry)
XX
DE Human cDNA encoding clone SKMUS20026340.
XX
KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW tissue regeneration; cell regeneration; membrane protein;
KW signal transduction-related protein; transcription-related protein;
KW osteoporosis; neurological disease; cancer; tumour.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT Location/Qualifiers
FT 57..1202
FT /*tag= a
FT /product= "Clone SKMUS20026340 protein"
XX
PN EPI308459-A2.
XX
PD 07-MAY-2003.
XX
PF 28-MAR-2002; 2002EP-00007401.
XX
PR 05-NOV-2001; 2001JP-00379298.
PR 25-JAN-2002; 2002US-00350978.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-450961/43.
DR P-PSDB; ADB65015.
XX
PT New polynucleotides and polypeptides, useful for developing a diagnostic
PT marker or medicines for regulation of their expression and activity, or
PT as targets of gene therapy.
XX
PS Claim 1; Page; 222pp; English.

XX The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesising the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a cDNA of the invention. Note: Some of the sequence
CC data for this patent is not represented in the printed specification, but
CC is based on sequence information supplied by the European Patent Office.

Sequence 1322 BP: 399 A; 318 C; 372 G; 233 T; 0 U; 0 Other; XX SQ

Query Match	99.9%	Score 1144.4	DB 9	Length 1322
Best Local Similarity	99.9%	Pred. No. 0		
Matches 1145	Conservative 0	Mismatches 1	Indels 0	Gaps 0
Qy	1	ATGGCTCCGGCGAAAAAGAAAGCGGGAGGCGCCAGCCAGAGACGCGCCCTCCGGAAGATA	60	
Db	57	ATGCTCCGGCGAAAAAGAAAGTGGGAGGCGCCAGCCAGAGCGCCCTCCGGAAGATA	116	
Qy	61	CGCAGCCACCTTGTCATCAGCTTGGCCGAGGTTGGCAGCAGTGGCGGAATGAGAAC	120	
Db	117	CGCAGCCACCTTGTCATCAGCTTGGCCGAGGTTGGCAGCAGTGGCGGAATGAGAAC	176	
Qy	121	AGCATCAGCAGGCGCCAGGAGCCCTACAGGCTGGCTGCCGGAGGAGCCAGGATCACT	180	
Db	177	AGCATCAGCAGGCGCCAGGAGCCCTACAGGCTGGCTGCCGGAGGAGCCAGGATCACT	236	
Qy	181	CAAGCTCTAAACNCATCACACCCCTACTTTCACACCAAGAAGCTCAGAGTGCCCAAG	240	
Db	237	CAAGCTCTAAACNCATCACACCCCTACTTTCACACCAAGAAGCTCAGAGTGCCCAAG	296	
Qy	241	TGCCACCCCGCTGCCAGAGGACATGGAGTGGACAAAGCTCAGAGAAAGCCCTGAG	300	
Db	297	TGCCACCCCGCTGCCAGAGGACATGGAGTGGACAAAGCTCAGAGAAAGCCCTGAG	356	
Qy	301	GTTTCTCATCAAAAGAAAGAGGTGTCCTCAAAACGGTGGTCAGCAAGACTTACGAGAGA	360	
Db	357	GTTTCTCATCAAAAGAAAGAGGTGTCCTCAAAACGGTGGTCAGCAAGACTTACGAGAGA	416	
Qy	361	GGAGGGAGCGTGAGCCACCTCAGCCACAGGTACGAGAGGATGCTGGTGCTTGAACCT	420	
Db	417	GGAGGGAGCGTGAGCCACCTCAGCCACAGGTACGAGAGGATGCTGGTGCTTGAACCT	476	
Qy	421	GGGCAGCCAGAAATGACATTGACAGATTCCTCCACAGCCACGGTCCCCCAACGGGAGG	480	
Db	477	GGGCAGCCAGAAATGACATTGACAGATTCCTCCACAGCCACGGTCCCCCAACGGGAGG	536	
Qy	481	AGAAATGTGCCAACCTGGTGCTGAGCTAACCAAGGGCTGGAGAGTGATGAGCAGGAG	540	
Db	537	AGAAATGTGCCAACCTGGTGCTGAGCTAACCAAGGGCTGGAGAGTGATGAGCAGGAG	596	
Qy	541	GAGCCACATGGAGGAGTCAGCGTAGACACAGAGGACAGGGCTATCGAGGAGAGCT	600	
Db	597	GAGCCACATGGAGGAGTCAGCGTAGACACAGAGGACAGGGCTATCGAGGAGAGCT	656	
Qy	601	GAGGAGAGGCCCGCAGCAGGATGAGGTGCAGGTGGCTGTGTGTCAAGATCAAGCGCCCTTG	660	

PT biodiversity.

Claim 1; SEQ ID NO 27986; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of states expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published/pct/sequences

Sequence 1373 BP; 396 A; 330 C; 408 G; 239 T; 0 U; 0 Other;

very Match 82.8%; Score 948.4; DB 5; Length 1373;

Best Local Similarity 86.0%; Pred. No. 1e-259;

Matches 1145; Conservative	0; Mismatches	1; Indels	186; Gaps	1;
----------------------------	---------------	-----------	-----------	----

Qy 1 ATGGCTCCGGCGAAAGGAAAGCGGGAGGGCCAGCCAGAGCGCCCTCCGGAAGATA 60

Db
42 ATGGCTCCGGCGGAAAGGAAAGCGGGGAGGGCCAGCCAGAGCGCCCTCCGGAAGATA 101

61 CGCACAGCCACCCCTGGTCATCAGCTTGGCCCCGAGGTTGGCAGCAGTGGCGGAATGAGAAC 120

Db 102 CGCACAGCCACCCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGCGCGAATGAGAAC 161

121 AGCATCAGGCAGGCTCAGGAGCCCTACAGGCTGGCTCCGGGAGGGACCCAGGACTCACCT 180

162 AGCATCAGGCAGGCCAGGAGCCTACAGCTGGTGCCTGGAGGACCCAGGACTCACCT 221

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

[illegible][illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

7

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

UU 4U2 GGAAGGACCCACAGGCACAGGTACCGAAGAAGGATGCTGGTGTGCTTGAACTT 408

421 GGGCAGCCAGAGAAAGACATATGACAGAAATCCCTCCACAGCCACGGCTCCCTCAACGCGGAGG 480

Db 462 GGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCCCCAACGGGAGG 521

QY 481 AGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCAGGAG 540

Db 522 AGAAAATGTGTCCCAACCTGGTGTCTGAGCTAACCAAGGCTGGAGAGTGATGGAGCAGGAG 581

QY 541 GAGCCACATGGAGGAGTGACAGCGTAGACACAGAGGACAGCGGCTATGGAGGAGAGGCT 600

D_b 582 GAGCCACATGGAGGAGTGACAGCGTAGACACAGAGGACAGCGGCTATGGAGGAGAGGCT 641

PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488901/53.
 XX
 XX Human genome-derived single exon nucleic acid probes useful for analyzing
 XX gene expression in human cervical epithelial cells.
 XX
 XX Claim 25; SEQ ID NO 7744; 487bp; English.
 XX
 XX The present invention relates to human single exon nucleic acid probes
 XX (SENPs). The present sequence is one such probe. The SENPs are derived
 XX from human Hela cells. The SENPs can be used to produce a single exon
 XX microarray, which can be used for measuring human gene expression in a
 XX sample derived from human cervical epithelial cells. By measuring gene
 XX expression, the probes are therefore useful in grading and/or staging of
 XX diseases of the cervix, notably cervical cancer. Note: The sequence data
 XX for this patent did not form part of the printed specification, but was
 XX obtained in electronic format directly from WIPO at
 XX ftp.wipo.int/pub/published_pct_sequences
 XX
 XX SQ Sequence 537 BP; 82 A; 161 C; 142 G; 152 T; 0 U; 0 Other;
 Query Match 43.6%; Score 499.4; DB 4; Length 537;
 Best Local Similarity 99.8%; Pred. No. 7e-132;
 Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 169 CAGGACTCCTCAAGCTCTTAACCAATCACACCCCTACTTCCACACCAAGAGCTCAG 228
 DB 537 CAGGACTCCTCAAGCTCTTAACCAATCACACCCCTACTTCCACACCAAGAGCTCAG 478
 QY 229 AGTGCCCAAGTCGCCACCCCGCTGCCAGAGGACATGGAGATGGACAAAGCTCAGAG 288
 DB 477 AGTGCCCAAGTCGCCACCCCGCTGCCAGAGGACATGGAGATGGACAAAGCTCAGAG 418
 QY 289 AAGGCCCTGAGGTTTCTCACATCAAAAGAGAGGTTGCCAAACCGTGGTCAGCAG 348
 DB 417 AAGGCCCTGAGGTTTCTCACATCAAAAGAGAGGTTGCCAAACCGTGGTCAGCAG 358
 QY 349 ACTTAGCAGAGAGGGGACGTGAGCCACCTCAGCCACGATACGAGAGGAGTCTGCT 408
 DB 357 ACTTAGCAGAGAGGGGACGTGAGCCACCTCAGCCACGATACGAGAGGAGTCTGCT 298
 QY 409 GTGCTTGAACCTGGGAGGAGGATGACATGACAGATGACATGACAGATCTCCACAGCAGGCTCC 468
 DB 297 GTGCTTGAACCTGGGAGGAGGATGACATGACAGATGACATGACAGATCTCCACAGCAGGCTCC 238
 QY 469 CCAACCGGAGGAGAGAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGTGGAGAGTG 528
 DB 237 CCAACCGGAGGAGAGAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGTGGAGAGTG 178
 QY 529 ATGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 588
 DB 177 ATGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 118
 QY 589 GGAGGAGAGGCTGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATC 648
 DB 117 GGAGGAGAGGCTGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATC 58
 QY 649 AAGGCCCTTGGCCCTCCAG 669
 DB 57 AAGGCCCTTGGCCCTCCAG 37

RESULT 4
 ABR62763/C
 ID ABR62763 standard; DNA; 537 BP.
 XX
 AC ABR62763;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human foetal liver single exon nucleic acid probe #11068.
 XX
 XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US000669.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-483447/52.
 XX
 XX Human genome-derived single exon nucleic acid probes useful for analyzing
 XX gene expression in human fetal liver.
 XX
 XX Claim 1; SEQ ID NO 11068; 639bp + Sequence Listing; English.
 XX
 XX The invention relates to a single exon nucleic acid probe for measuring
 XX human gene expression in a sample derived from human foetal liver. The
 XX single exon nucleic acid probes may be used for predicting, measuring and
 XX displaying gene expression in samples derived from human fetal liver. The
 XX present sequence is a single exon nucleic acid probe of the invention.
 XX Note: The sequence data for this patent did not form part of the printed
 XX specification, but was obtained in electronic format directly from WIPO
 XX at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX SQ Sequence 537 BP; 82 A; 161 C; 142 G; 152 T; 0 U; 0 Other;
 Query Match 43.6%; Score 499.4; DB 4; Length 537;
 Best Local Similarity 99.8%; Pred. No. 7e-132;
 Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 169 CAGGACTCCTCAAGCTCTTAACCAATCACACCCCTACTTCCACACCAAGAGCTCAG 228
 DB 537 CAGGACTCCTCAAGCTCTTAACCAATCACACCCCTACTTCCACACCAAGAGCTCAG 478
 QY 229 AGTGCCCAAGTCGCCACCCCGCTGCCAGAGGACATGGAGATGGACAAAGCTCAGAG 288
 DB 477 AGTGCCCAAGTCGCCACCCCGCTGCCAGAGGACATGGAGATGGACAAAGCTCAGAG 418
 QY 289 AAGGCCCTGAGGTTTCTCACATCAAAAGAGAGGTTGCCAAACCGTGGTCAGCAG 348
 DB 417 AAGGCCCTGAGGTTTCTCACATCAAAAGAGAGGTTGCCAAACCGTGGTCAGCAG 358
 QY 349 ACTTAGCAGAGAGGGGACGTGAGCCACCTCAGCCACGATACGAGAGGAGTCTGCT 408
 DB 357 ACTTAGCAGAGAGGGGACGTGAGCCACCTCAGCCACGATACGAGAGGAGTCTGCT 298
 QY 409 GTGCTTGAACCTGGGAGGAGGATGACATGACAGATGACATGACAGATCTCCACAGCAGGCTCC 469
 DB 297 GTGCTTGAACCTGGGAGGAGGATGACATGACAGATGACATGACAGATCTCCACAGCAGGCTCC 238

PS Claim 1; SEQ ID NO 8525; 530pp; English.
 XX The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart. The
 CC present sequence is one such probe. The probes may be used for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from the human heart via microarrays. By measuring gene expression, the
 CC probes are useful for predicting, diagnosing, grading, staging,
 CC monitoring and prognosing diseases of the human heart and vascular system
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
 CC congenital heart disease. Note: The sequence data for this patent did not
 CC form part of the printed specification, but was obtained in electronic
 CC format directly from WIPO at fep.wipo.int/pub/published_pct_sequences
 XX
 XX
 SQ Sequence 537 BP; 82 A; 161 C; 142 G; 152 T; 0 U; 0 Other;

Query Match 43.6%; Score 499.4; DB 4; Length 537;
 Best Local Similarity 99.8%; Pred. No. 7e-132;
 Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 169 CAGGACTCAGCTCAAGCTCTAAACCAATCACACCCCTTACTTACACACCAAGGCTCAG 228
 Db 537 CAGGACTCAGCTCAAGCTCTAAACCAATCACACCCCTTACTTACACACCAAGGCTCAG 478
 QY 229 AGTGCCCAAGTCGCCCGCCGCTCCAGAGGACATGGAGTGGCAAGGCTCAGAG 288
 Db 477 AGTGCCCAAGTCGCCCGCCGCTCCAGAGGACATGGAGTGGCAAGGCTCAGAG 418
 QY 289 AAAGCCCTGAGGTTTCTCACATCAAAAGAAAGAGGTTGTCACCAACGGTGGTCAAG 348
 Db 417 AAAGCCCTGAGGTTTCTCACATCAAAAGAAAGAGGTTGTCACCAACGGTGGTCAAG 358
 QY 349 ACTTAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 408
 Db 357 ACTTAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 298
 QY 409 GTGCTTGAACCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 468
 Db 297 GTGCTTGAACCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 238
 QY 469 CCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 528
 Db 237 CCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 178
 QY 529 ATGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 588
 Db 177 ATGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 118
 QY 589 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 648
 Db 117 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 58
 QY 649 AAGCGCCCTTGGCCCTCCAG 669
 Db 57 AAGCGCCCTTGGCCCTCCAG 37

RESULT 7
 AK36965/c
 ID AK36965 standard; DNA; 537 BP.
 XX
 AC AK36965;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human bone marrow expressed single exon probe SEQ ID NO: 11522.
 XX
 KW Human; bone marrow expressed exon; gene expression analysis; probe;
 XX Microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 XX Homo sapiens.
 OS
 XX
 FN WO200157276-A2.

XX 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US000668.
 XX 04-FEB-2000; 2000US-0180312P.
 XX 26-MAY-2000; 2000US-0207456P.
 PR 03-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-489900/53.
 DR Human genome-derived single exon nucleic acid probes useful for analyzing
 XX gene expression in human bone marrow.
 XX Example 4; SEQ ID NO 11522; 658pp + Sequence Listing; English.
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention
 XX
 SQ Sequence 537 BP; 82 A; 161 C; 142 G; 152 T; 0 U; 0 Other;
 Query Match 43.6%; Score 499.4; DB 4; Length 537;
 Best Local Similarity 99.8%; Pred. No. 7e-132;
 Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 169 CAGGACTCAGCTCAAGCTCTAAACCAATCACACCCCTTACTTACACACCAAGGCTCAG 228
 Db 537 CAGGACTCAGCTCAAGCTCTAAACCAATCACACCCCTTACTTACACACCAAGGCTCAG 478
 QY 229 AGTGCCCAAGTCGCCCGCCGCTCCAGAGGACATGGAGTGGCAAGGCTCAGAG 288
 Db 477 AGTGCCCAAGTCGCCCGCCGCTCCAGAGGACATGGAGTGGCAAGGCTCAGAG 418
 QY 289 AAAGCCCTGAGGTTTCTCACATCAAAAGAAAGAGGTTGTCACCAACGGTGGTCAAG 348
 Db 417 AAAGCCCTGAGGTTTCTCACATCAAAAGAAAGAGGTTGTCACCAACGGTGGTCAAG 358
 QY 349 ACTTAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 408
 Db 357 ACTTAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 298
 QY 409 GTGCTTGAACCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 468
 Db 297 GTGCTTGAACCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 238
 QY 469 CCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 528
 Db 237 CCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 178
 QY 529 ATGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 588
 Db 177 ATGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 118
 QY 589 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 648
 Db 117 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 58
 QY 649 AAGCGCCCTTGGCCCTCCAG 669
 Db 57 AAGCGCCCTTGGCCCTCCAG 37

```

RESULT 8
AAK11153/c
ID AAK11153 standard; DNA; 537 BP.
AC AAK11153;
XX
DT 05-NOV-2001 (first entry)
DE Human brain expressed single exon probe SEQ ID NO: 11144.
XX
KW Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
KW ss.
XX
OS Homo sapiens.
XX
FN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000667.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-483446/52.
XX
Single exon nucleic acid probes for analyzing gene expression in human
brains.
XX
Example 4; SEQ ID NO 11144; 650pp + Sequence Listing; English.
XX
The present invention provides a number of single exon nucleic acid
probes which are derived from genomic sequences expressed in the human
brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention
XX
SQ Sequence 537 BP; 82 A; 161 C; 142 G; 152 T; 0 U; 0 Other;
Query Match 43.6%; Score 499.4; DB 4; Length 537;
Best Local Similarity 99.8%; Pred. No. 7e-132;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 169 CAGGACTCAGCTCAAGTCTCTTAACGATCAGACCCCTTCTCAGCAGGAGGAGGATGCTGCTG 228
DB |||||
QY 537 CAGGACTCAGCTCAAGTCTCTTAACGATCAGACCCCTTCTCAGCAGGAGGAGGATGCTGCTG 478
DB |||||
QY 229 AGTCCGCCAAAGTCGCCACCCCGCTGCGCAGAGGACATGGAGATGGACAAAGCTCAGAG 288
DB |||||
QY 477 AGTCCGCCAAAGTCGCCACCCCGCTGCGCAGAGGACATGGAGATGGACAAAGCTCAGAG 418
DB |||||
QY 289 AAAGCCCTGAGGTTTCTCATCAAAAGAGAGGTGTCAGGAGGAGGAGGAGGAGGAGGAGGAG 348
DB |||||
QY 417 AAAGCCCTGAGGTTTCTCATCAAAAGAGAGGTGTCAGGAGGAGGAGGAGGAGGAGGAGGAG 358
DB |||||
QY 349 ACTTACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 408
DB |||||
QY 357 ACTTATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 298
DB |||||
QY 409 GTGCTTGAACCTGGGCGAGCAGAGATGACATTGACAGATCTCCACAGCCACGGCTCC 468
DB |||||
DB 297 GTGCTTGAACCTGGGCGAGCAGAGATGACATTGACAGATCTCCACAGCCACGGCTCC 238
QY 469 CCAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 528
DB |||||
QY 237 CCAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 178
QY 529 ATGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 588
DB |||||
DB 177 ATGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 118
QY 589 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATC 648
DB |||||
QY 117 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATC 58
QY 649 AAGCGCCCTTGGCCCTCCAG 669
DB 57 AAGCGCCCTTGGCCCTCCAG 37
RESULT 9
ABS36644/c
ID ABS36644 standard; DNA; 537 BP.
XX
AC ABS36644;
XX
DT 25-FEB-2003 (first entry)
DE Human liver single exon probe, SEQ ID NO 11634.
XX
KW Human; single exon nucleic acid probe; liver; cirrhosis;
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW coronary heart disease; ss.
XX
OS Homo sapiens.
XX
FN WO200157273-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000664.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-488898/53.
XX
Human genome-derived single exon nucleic acid probes useful for analyzing
gene expression in human adult liver.
XX
Claim 1; SEQ ID NO 11634; 658pp; English.
XX
The invention relates to a single exon nucleic acid probe (SENP) (I) for
measuring human gene expression in a sample derived from human adult
liver, comprising one of 13109 defined nucleotide sequences given in the
specification (or complements/fragments). The probe hybridises at high
stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABS25011-ABS51005 represent human
CC liver single exon nucleic acid probes of the invention. Note: The
CC sequence information for this patent does not appear in the printed

```



```
PN WO200157277-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000669.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human fetal liver.
XX Claim 4; SEQ ID NO 23580; 639pp + Sequence Listing; English.
XX The invention relates to a single exon nucleic acid probe for measuring
XX human gene expression in a sample derived from human fetal liver. The
XX single exon nucleic acid probes may be used for predicting, measuring and
XX displaying gene expression in samples derived from human fetal liver. The
XX present sequence is a single exon nucleic acid probe of the invention.
XX Note: the sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 365 BP; 49 A; 126 C; 78 G; 112 T; 0 U; 0 Other;
XX
Query Match 31.7%; Score 363.4; DB 4; Length 365;
Best Local Similarity 99.7%; Pred. No. 3.5e-93;
Matches 364; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 260 AAGGACATGGAGATGACAAAGCTCAGAGAAAGCCCTGAGGTTTCTCACATCAAAAGA 319
DB 365 AAGGACATGGAGATGACAAAGCTCAGAGAAAGCCCTGAGGTTTCTCACATCAAAAGA 306
QY 320 AAGAGTGTCCTCCAAACCGTGTCTCAGCAGACTTACGAGAGAGGGGAGCTGAGCCACC 379
DB 305 AAGAGTGTCCTCCAAACCGTGTCTCAGCAGACTTATGAGAGAGAGGGGAGCTGAGCCACC 246
QY 380 TCAGCCACAGGTACGAGAGGGATGCTGTGTCTTGAACCTGGGACCCAGAGAAATGACA 439
DB 245 TCAGCCACAGGTACGAGAGGGATGCTGTGTCTTGAACCTGGGACCCAGAGAAATGACA 186
QY 440 TTGACAGATCTCCAAACCGTGTCTCAGCAGACTTACGAGAGAGGGGAGCTGAGCCACC 499
DB 185 TTGACAGATCTCCAAACCGTGTCTCAGCAGACTTATGAGAGAGAGGGGAGCTGAGCCACC 126
QY 500 TGTCTGAGCTAACCAAGGCTGGAGAGTGTGAGAGCAGGAGAGCCACATGAGGAGTG 559
DB 125 TGTCTGAGCTAACCAAGGCTGGAGAGTGTGAGAGCAGGAGAGCCACATGAGGAGTG 66
QY 560 ACAGCGTAGACACAGAGGACAGCGGCTATGGAGGAGGCTGAGAGAGGCGCCAGCAGG 619
DB 65 ACAGCGTAGACACAGAGGACAGCGGCTATGGAGGAGGCTGAGAGAGGCGCCAGCAGG 6
QY 620 ATGGA 624
DB 5 ATGGA 1
RESULT 12
AAI55838/c
ID AAI55838 standard; DNA; 365 BP.
XX
```

```
AAI55838;
XX 17-OCT-2001 (first entry)
XX Probe #24524 used to measure gene expression in human placenta sample.
XX Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX Homo sapiens.
XX WO200157272-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000663.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX Claim 25; SEQ ID NO 24524; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders
XX
XX Sequence 365 BP; 49 A; 126 C; 78 G; 112 T; 0 U; 0 Other;
XX
Query Match 31.7%; Score 363.4; DB 4; Length 365;
Best Local Similarity 99.7%; Pred. No. 3.5e-93;
Matches 364; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 260 AAGGACATGGAGATGACAAAGCTCAGAGAAAGCCCTGAGGTTTCTCACATCAAAAGA 319
DB 365 AAGGACATGGAGATGACAAAGCTCAGAGAAAGCCCTGAGGTTTCTCACATCAAAAGA 306
QY 320 AAGAGTGTCCTCCAAACCGTGTCTCAGCAGACTTACGAGAGAGGGGAGCTGAGCCACC 379
DB 305 AAGAGTGTCCTCCAAACCGTGTCTCAGCAGACTTATGAGAGAGAGGGGAGCTGAGCCACC 246
QY 380 TCAGCCACAGGTACGAGAGGGATGCTGTGTCTTGAACCTGGGACCCAGAGAAATGACA 439
DB 245 TCAGCCACAGGTACGAGAGGGATGCTGTGTCTTGAACCTGGGACCCAGAGAAATGACA 186
QY 440 TTGACAGATCTCCAAACCGTGTCTCAGCAGACTTACGAGAGAGGGGAGCTGAGCCACC 499
DB 185 TTGACAGATCTCCAAACCGTGTCTCAGCAGACTTATGAGAGAGAGGGGAGCTGAGCCACC 126
QY 500 TGTCTGAGCTAACCAAGGCTGGAGAGTGTGAGAGCAGGAGAGCCACATGAGGAGTG 559
DB 125 TGTCTGAGCTAACCAAGGCTGGAGAGTGTGAGAGCAGGAGAGCCACATGAGGAGTG 66
QY 560 ACAGCGTAGACACAGAGGACAGCGGCTATGGAGGAGGCTGAGAGAGGCGCCAGCAGG 619
DB 65 ACAGCGTAGACACAGAGGACAGCGGCTATGGAGGAGGCTGAGAGAGGCGCCAGCAGG 6
QY 620 ATGGA 624
```

```

Db          5 ATGGA 1
|||||
RESULT 13
ABA39931/c
ID ABA39931 standard; DNA; 365 BP.
AC
AC ABA39931;
XX
XX
DT 23-JAN-2002 (first entry)
DE
DE Probe #18397 for gene expression analysis in human heart cell sample.
XX
XX Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
XX WO200157274-A2.
PN
XX
XX 09-AUG-2001.
ED
XX
XX 30-JAN-2001; 2001WO-US000666.
FF
XX
XX 04-FEB-2000; 2000US-0180312P.
PR
XX 26-MAY-2000; 2000US-0207456P.
PR
XX 30-JUN-2000; 2000US-00608408.
PR
XX 03-AUG-2000; 2000US-00632366.
PR
XX 21-SEP-2000; 2000US-0234687P.
PR
XX 27-SEP-2000; 2000US-0236359P.
PR
XX 04-OCT-2000; 2000US-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts.
XX
XX Claim 4; SEQ ID NO 18397; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease. Note: The sequence data for this patent did not
XX form part of the printed specification, but was obtained in electronic
XX format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 365 BP; 49 A; 126 C; 78 G; 112 T; 0 U; 0 Other;
XX
XX Query Match 31.7%; Score 363.4; DB 4; Length 365;
XX Best Local Similarity 99.7%; Pred. No. 3.5e-93;
XX Matches 364; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 260 AAGACATGGAGTGGACAAAGCTCAGAGAGCCCTGAGGTTCTCATCAAAAGA 319
XX |||||||
XX 365 AAGCATGGAGTGGACAAAGCTCAGAGAGCCCTGAGGTTCTCATCAAAAGA 306
XX |||||||
XX 320 AAGAGGTGTCCAAACCGTGTGTCAGCAAGACTTACGAGAGAGGAGCGTGGCCACC 379
XX |||||||
XX 305 AAGAGGTGTCCAAACCGTGTGTCAGCAAGACTTATGAGAGAGGAGCGTGGCCACC 246
XX |||||||
XX 380 TCAGCCACAGTACGAGAGGAGTCTGGTGTCTTGAACCTGGGCGCCAGCAATGACA 439
XX |||||||

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: May 4, 2004, 08:49:08 ; Search time 18 Seconds
(without alignments)
6630.266 Million cell updates/sec

Title: US-10-644-659A-1

Perfect score: 2058
Sequence: 1 atggctccggcgaaagga.....tgattacgctactcaagtga 1146

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cgn2.1/USPTO_spool_p/US10644659/runat_04052004_084902_28510/app_query.fasta_1.1287
-DB=SwissProt_42 -QPM=fastan -SUFFIX=n2p.rap -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10644659 @CNC 1.1.16 @runat_04052004_084902_28510 -NCPU=6 -ICPU=3
-NO MWAP -LARGEQUERY -NEG_SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	130.5	6.3	1739	1 CHD2_HUMAN	O14647 homo sapien
2	123	6.0	838	1 GLT4_WHEAT	P08489 triticum ae
3	122	5.9	887	1 NED4_RAT	O62940 rattus norv
4	121	5.9	2805	1 MAF2_HUMAN	P78553 homo sapien
5	120	5.8	839	1 GLT5_WHEAT	P10388 triticum ae
6	120	5.8	1182	1 HAIR_MOUSE	Q61645 mus musculu
7	119.5	5.8	458	1 YNE1_CABEL	P30640 caenorhabdi
8	118.5	5.8	1067	1 MICA_HUMAN	Q8tdz2 homo sapien
9	116.5	5.7	1001	1 IF2_SYNY3	P72689 synecocyst
10	115.5	5.6	2472	1 NCR2_MOUSE	Q9wu42 mus musculu
11	115	5.6	646	1 SGL_BOVIN	P23389 bos taurus
12	114	5.5	1275	1 TCOF_DROME	P19334 drosophila
13	114	5.5	1411	1 XPC_HUMAN	Q3428 homo sapien
14	113.5	5.5	1293	1 XPC_DROME	Q34595 drosophila
15	113	5.5	350	1 VSK1_CHICK	Q1a12 gallus gall
16	113	5.5	1125	1 IF2_PROMM	Q7v5m4 prochloroco
17	113	5.5	1593	1 AT12_HUMAN	P88397 homo sapien
18	113	5.4	2003	1 NTC4_HUMAN	Q99466 homo sapien

ALIGNMENTS

RESULT 1

ID	CHD2_HUMAN	STANDARD	PRT	1739 AA
AC	O14647			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Chromodomain-helicase-DNA-binding protein 2 (CHD-2).			
GN	CHD2.			
OS	homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97470991; PubMed=9326634;			
RA	Woodage T., Basrai M.A., Baxevanis A.D., Hieter P., Collins F.S.;			
RT	"Characterization of the CHD family of proteins."			
RL	Proc. Natl. Acad. Sci. U.S.A. 94:11472-11477(1997);			
CC	-!- FUNCTION: Sequence-selective DNA-binding protein (By similarity).			
CC	-!- SUBCELLULAR LOCATION: Nuclear (By similarity).			
CC	-!- SIMILARITY: Belongs to the SNF2/SAD54 helicase family.			
CC	-!- SIMILARITY: Contains 2 chromo domains.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	EMBL; AF006514; AAB87382.1; -			
DR	HSSP; P23197; 1A0P.			
DR	Genew; HGNC:1917; CHD2.			
DR	MIM; 602113; -			
DR	GO; GO:0004003; F:ATP dependent DNA helicase activity; TAS.			
DR	GO; GO:0003677; F:DNA binding; TAS.			
DR	GO; GO:0007001; P:chromosome organization and biogenesis (gen. . . ; TAS.			
DR	GO; GO:0006357; P:regulation of transcription from Pol II pro. . . ; TAS.			
DR	InterPro; IPR000953; Chromo.			
DR	InterPro; IPR001410; DEAD.			

DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR000330; SNE2_N.
 DR Pfam; PF00385; chromo; 2.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00176; SNE2_N; 1.
 DR SMART; SM00298; CHROMO; 2.
 DR SMART; SM00487; DEXDs; 1.
 DR SMART; SM00490; HELIC; 1.
 DR PROSITE; PS00598; CHROMO_1; 2.
 DR PROSITE; PS00013; CHROMO_2; 2.
 DR PROSITE; PS00690; DEAH ATP HELICASE; FALSE NEG.
 KW Hydrolyase; Helicase; Nuclear protein; ATP-Binding; DNA-binding;
 KW Repeat.
 FT DOMAIN 1 74 SER-RICH.
 FT DOMAIN 121 139 SER-RICH.
 FT DOMAIN 261 353 CHROMO 1.
 FT DOMAIN 378 456 CHROMO 2.
 FT NP_BIND 509 516 ATP (POTENTIAL).
 FT SITE 617 620 DEAH BOX.
 SQ SEQUENCE 1739 AA; 200560 MW; D62089C25EBA40F CRC64;
 Alignment Scores:
 Pred. No.: 0.163 Length: 1739
 Score: 130.50 Matches: 63
 Percent Similarity: 37.10% Conservative: 42
 Best Local Similarity: 22.26% Mismatches: 123
 Query Match: 6.34% Indels: 55
 DB: 1 Gaps: 10

US-10-644-659A-1 (1-1146) x CHD2_HUMAN (1-1739)
 QY 109 GCGAATGAGAACAGCATCAGGCGGCGGAGGAGGCTACAGGCTGGCTGCGGGAGGACC 168
 Db 17 SerAsnAlaSerSerHisserAlaSerGluGluAlaSerGlySerAspSerGlySerGln 36
 QY 169 CAGGACTACCTCAGCTCCTTAACCAATCACACCCCTACTTACAC---CAGAAAGCT 225
 Db 37 SerGluSerGluGluGlnGlySer-----AspProGlySerGlyHisGlySerGluSer 53
 QY 226 CAGAGTGCCCAAGTCGCCACCCCGCTGCCAGAGGACATGAGATGGAGCAAGCTCA 285
 Db 54 AsnSerSerSerGluSerGluSerGluSerGlnSerGluSerGluSerGluSerAlaGlySer 73
 QY 286 GAGAAAGCCCTGAGTCTTCATCAATCAAAAGAAAGAGGTGTCGAAACCGGTGGTCAGC 345
 Db 74 LysSerGlnProValLeuProGluAlaLysGluLysProAlaSerLysLys----- 90
 QY 346 AAGACTTACAGAGAGGAGGAGGAGCGGACGTGACCCACCTCAGCCACAGGTACGAGGATGCT 405
 Db 91 -----GluArgIleAlaAspValLysLysMetTrpGluGluTyrProAspValTyr 107
 QY 406 GTGTGCTTGAACCTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 453
 Db 108 GlyValArgArgSerAsnArgSerArgGlnGluProSerArgPheAsnIleLysGluGlu 127
 QY 454 -----CACAGCCACCGCTCCCAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 501
 Db 128 AlaSerSerGlySerGluSerGlySerProLysAspArg----- 140
 QY 502 TCTGAGTAAACCAAGGCTGGAGAGTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 561
 Db 141 -----GlyGlnArgGlnLeuLysLysGlnGluLysTrpLysGlnGluPro 155
 QY 562 AGCGTAGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 621
 Db 156 SerGluAspGluGlnGlnGlnGlyThrSerAlaGluSerGlu-----ProGluGln--- 172
 QY 622 GGAGTGCAGGTGCTGTGTCAGGATCAAGCGCCCTTGCCTCCAGGTAAACAGATTT 681
 Db 173 -----LysLysValLysAlaArgProValPro-----ArgArgThr 185
 QY 682 ACAGAGAACTCACTGCAAGGCCCAACAGAAATATAGCCAGTGGCGCAACTTGAAGGG 741

Db 186 ValProLysProArgValLysLysGlnProLys-----ThrGlnArgGlyLysArg 202
 QY 742 AGATGGCAGCAGTGGCTGATGAA----- 765
 Db 203 LysLysGlnAspSerSerAspGluAspAspAspAspGluAlaProLysArgGlnThr 222
 QY 766 CACATACATCCAGAGAGCTCACTCTTCAGTGTAGAGTTGATTACGAGCTGGCCATG 825
 Db 223 ArgArgArgAlaAlaLysAsnValSerLysGluAspAspAspPheGluThrAspSer 242
 QY 826 TCCACCCGCTACACAAAGGAGATGAGGCTATGCGCCGCCCCAAGAGGAGCAACAACT 885
 Db 243 AspaAspLeuIleGluMetThrGlyGluGlyValAspGluGlnGlnAspAsnSerGluThr 262
 QY 886 GCTGAAGG 894
 Db 263 IleGluLys 265

RESULT 2
 GLT4 WHEAT
 ID GLT4 WHEAT STANDARD; PRT; 838 AA.
 AC P08489;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Glutinin, high molecular weight subunit Pw212 precursor.
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
 OC Triticeae; Triticum.
 OX NCBI_TaxID=4565;
 RN [1] SEQUENCE FROM N.A.
 RC STRAIN=cv. Yamhill;
 RX MEDLINE=86093674; PubMed=3001648;
 RA Sugiyama T., Rafalski A., Peterson D., Soll D.G.;
 RT "A wheat HMW glutenin subunit gene reveals a highly repeated
 RT structure";
 RL Nucleic Acids Res. 13:8729-8737(1985).
 CC -!- FUNCTION: Glutens are the high molecular weight seed storage
 CC protein of wheat endosperm. Thought to be responsible for the
 CC visco-elastic property of wheat dough.
 CC -!- SUBUNIT: Disulfide-bridge linked aggregates.
 CC -!- MISCELLANEOUS: Glutens are coded by several genes on each of the
 CC group 1 chromosomes of wheat.
 CC -!- MISCELLANEOUS: The mature protein is characterized by a large
 CC number of well preserved repeats of the two motifs: GQQPGQ and
 CC GQQPGQGGQGYPT.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X03346; CAA27052.1; -
 DR FIR; A24107; SEWTHW.
 DR InterPro; IPR003612; AAI.
 DR InterPro; IPR001419; Glutinin.
 DR Pfam; PF03157; Glutinin hmw; 1.
 DR PRINTS; PR00210; GLUTENIN.
 DR Seed storage protein; Repeat; Multigene family; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 838 GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT
 FT PW212.
 FT DOMAIN 130 799 REPEATS.
 SQ SEQUENCE 838 AA; 89174 MW; 71D715B7BDF0722D CRC64;
 Alignment Scores:
 Pred. No.: 0.492 Length: 838
 Score: 123.00 Matches: 75

FT BINDING 854 854 UBIQUITIN (BY SIMILARITY).
SQ SEQUENCE 887 AA; 102394 MW; D74B1097688CD9A1 CRC64;

Alignment Scores:
Pred. No.: 0.579 Length: 887
Score: 122.00 Matches: 86
Percent Similarity: 35.85% Conservations: 66
Best Local Similarity: 20.28% Mismatches: 128
Query Match: 5.93% Indels: 144
DB: 1 Gaps: 21

US-10-644-659A-1 (1-1146) x NED4_RAT (1-887)

QY 34 CAGCCAGAGCCCTCCGAGATAGCAGCAGCAGCCCTGCTCATCAGCTTGGCCCA 93
Db 107 ProMetSerGlyValLeuThrSerValGlnThrLysThrLysLysSerLeuAsnPro 126
QY 94 GGTGGCAGCAGC----- 105
Db 127 LysTrpAsnGluGluLeuPheArgValLeuProGlnGlnHisArgIleLeuPheGlu 146
QY 106 -----TGGCCGAATGAGAACAGCATCAGCAGCAGCCCGCCAGGAGCCTACA 147
Db 147 ValPheAspGluAsnArgLeuThrArgAspPheLeuGlyGlnValAspValProLeu 166
QY 148 GGCTGGCTGCCGAGGAGCAGCAGCTCACTCAAGCTCCTAAACCAATCACA----- 201
Db 167 TyrProLeuPro-----ThrGluAsnProArgMetGluArgProTyrThrPheLys 183
QY 202 -----CCCCCTACTTCACACAGAAAGCTCAGAGTGCCTCCCAAGTGCACCC 249
Db 184 AspPheValLeuHisProArgSerHis-----LysSerArgValLysGlyTyrLeu 200
QY 250 CGC-----CTGCCAGAGGACATGAGATGAGCAGCAAGCTCAGAGAAAGCC 294
Db 201 ArgLeuLysMetThrTyrLeuProLysAsnGlySerAspAspGluAsnAlaAspGlnAla 220
QY 295 CCTGAGGT-----TTCACATCAAA 315
Db 221 GluGluLeuGluProGlyTyrValValLeuAspGlnProAspAlaAlaThrHisLeuGln 240
QY 316 AAGAAA-----GAGGTGTCCAAAGCTGCTCAGC 345
Db 241 HisProGluProSerProLeuProGlyTyrGluGluArgGlnAspValLeuGly 260
QY 346 AAGATTACGAGAGAGGAGGAGCTGAGCCACTCAGCCAC-----AGTACGAGAG 399
Db 261 ArgThrTyr-----ValAsnHisGluSerArgThrThrGlnTyrLysArg 276
QY 400 GATGCTGTGCTTGAACCTGGGCGAGCCAGAGATGACAT----- 441
Db 277 ProSer-----ProGluAspAspLeuThrAspAspGluAsnGly 289
QY 442 GACAAATCTCCACAGCAGCCGCTCCCAAGCGGAGGAGAAATGTGCCAACTGGTG 501
Db 290 AspIleGlnLeuGlnAlaHisGlyAlaPheThrThrArgArgGlnLysSerGluAspVal 309
QY 502 TCT-----GAGCTAACAGGCTGGAGAGTGTGAGCAGCAGGAGCC 546
Db 310 AspGlyProAspAsnHisGluSerProGluAsnTrpGluIleValArgGluAspGluAsn 329
QY 547 ACATGAGAGGTGACAGCGTACACAGAGGACCGCTGAGGAGAGAGCTGAGGAG 606
Db 330 ThrIle-----TyrSerGlyGlnAlaValGln 338
QY 607 AGGCCCGCAGGATGAGTGCAGTGGCTGTGTCAGATCAAGCCGCTTGCCTCC 666
Db 339 SerProSerGlyHisProAspValGln-----ValArgLeuAlaGluLeuAspThr 357
QY 667 CAGGTAACAGATTACAGAGAACTCACTCAAGCC-----AACAGAAATATAGC 720
Db 358 ArgLeuThrMetTyrGlyAsnProAlaThrSerGlnProValThrSerAsnHis 377

QY 721 CCACTGGGCACTTGAAGGAGATGCGCAGCAGTGGCTGATGAACACATACATCCAG 780
Db 377 rSer-----ArgGlyGlySerSerGlnThrCysIlePheGluGlnProTh 393
QY 781 AAGCTCAATCCTTTTCAGTGAAGAGTTTATTACAGAGTGGCCATGTCCACCC----- 832
Db 393 rLeuProValLeu-----LeuProThrSerSerGlyLeuProGlyTyrG 409
QY 833 -----GCTACACAAAGAGATGAGGCTATGGCCGCC 867
Db 409 uGluLysGlnAspAspArgGlyArgSerTyrTyrValAspHisAsnSerLysThrTh 429
QY 868 AAGAAGAGGAAACCAACTGCTGAAAGGGCCCA-----AGCTGCTGAGGAGCAGATACAGG 924
Db 429 rTrpSerLysProThrMetGlnAspAspProArgSerLysIle-ProAlaHisLeuArg 449
QY 925 GA----- 926
Db 449 lYlYsThrProValAspSerAsnAspLeuGlyProLeuProProGlyTyrGluGluArgT 469
QY 927 -----AATGATGGACATGCTGCTTCAATATCTGCAATGGCTCGCCACAGAGATGCC 981
Db 469 hrHisThrAspGlyArgValPhePheIleAsnHisAsn-----IleLysLysThrGlnTrpG 488
QY 982 AAGATCCA 989
Db 488 luAspPro 490
RESULT 4
MAPA_HUMAN
ID MAPA_HUMAN STANDARD; PRT: 2805 AA.
AC P78559; Q95643; Q12973; Q15882; Q9UUT4;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Microtubule-associated protein 1A (MAP 1A) (proliferation-related
protein p80) [Contains: MAP1 light chain LC2].
GN MAP1A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97001161; PubMed=8812494;
RA Fink J.K., Jones S.M., Esposito C., Wilkowski J.;
RT "Human microtubule-associated protein 1A (MAP1A) gene: genomic
organization, cDNA sequence, and developmental- and tissue-specific
expression.";
RL Genomics 35:577-585(1996).
RN [2]
RP SEQUENCE OF 1-1825 FROM N.A.
RX TISSUE=Ovarian carcinoma;
RA Chen Z.C., Fadiel A., Naftolin F.;
RT "Identification of a novel protein (p80) in ovarian carcinoma cells.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 78-1687 FROM N.A.
RX TISSUE=Brain;
RA Ohtani K., Rutherford T., Sakamoto H., Naftolin F.;
RT "Microtubule associated protein 1A (MAP1A) in human brain - DNA
sequence and physiological role.";
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 134-419 FROM N.A.
RX TISSUE=Fetal muscle;
RA Chianilukhal N., Pasturaud P., Richard I., Auffray C.,
RA Beckmann J.S.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1607-1883 FROM N.A.
RX TISSUE=Brain;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1607-1883 FROM N.A.
RX MEDLINE=95356255; PubMed=7629894;

RA Fukuyama R., Rapoport S.I.;
 RT "Brain-specific expression of human microtubule-associated protein 1A
 RL (MAP1A) gene and its assignment to human chromosome 15.";
 CC J. Neurosci. Res. 40:820-825(1995).
 CC -/- FUNCTION: Structural protein involved in the filamentous cross-
 CC bridging between microtubules and other skeletal elements.
 CC -/- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
 CC with MAP1A and MAP1B proteins.
 CC -/- TISSUE SPECIFICITY: Brain.
 CC -/- DOMAIN: The basic region containing the repeats may be responsible
 CC for the binding of MAP1A to microtubules.
 CC -/- PTM: Various serine residues may be phosphorylated by cAMP kinase.
 CC -/- PTM: LC2 IS COEXRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED
 CC FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
 CC BOTH MAP1A AND MAP1B.
 CC -/- SIMILARITY: TO MAP1B.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).

DR EMBL; U38291; AAB41132.1; -;
 DR EMBL; U38292; AAB41133.1; -;
 DR EMBL; AF200415; AAF08305.2; -;
 DR EMBL; U80458; AAD00355.1; -;
 DR EMBL; 247038; CAA87104.1; -;
 DR EMBL; U14577; AAB81362.1; -;
 DR PIR; I38857; I38857.
 DR Genew; HGNC:6835; MAP1A.
 DR MIM; 600178; -;
 DR GO; GO:0005875; C:microtubule associated complex; TAS.
 KW Microtubule; Repeat; Phosphorylation.
 FT CHAIN 2490 2805
 FT DOMAIN 309 496
 FT LYS-RICH (BASIC).
 FT 9 X 3 AA REPEATS OF K-K-[DE].
 FT REPEAT 415 541
 FT REPEAT 415 417
 FT REPEAT 420 422
 FT REPEAT 427 429
 FT REPEAT 431 433
 FT REPEAT 436 438
 FT REPEAT 440 442
 FT REPEAT 444 446
 FT REPEAT 449 451
 FT REPEAT 459 541
 FT REPEAT 72 72
 FT CONFLICT 134 135
 FT CONFLICT 249 249
 FT CONFLICT 263 263
 FT CONFLICT 296 296
 FT CONFLICT 311 311
 FT CONFLICT 324 324
 FT CONFLICT 324 324
 FT CONFLICT 325 326
 FT CONFLICT 333 333
 FT CONFLICT 357 357
 FT CONFLICT 364 364
 FT CONFLICT 414 419
 FT CONFLICT 424 424
 FT CONFLICT 426 426
 FT CONFLICT 431 431
 FT CONFLICT 439 439
 FT CONFLICT 444 444
 FT CONFLICT 452 453
 FT CONFLICT 457 457
 FT CONFLICT 682 682
 FT CONFLICT 1025 1025
 FT CONFLICT 1303 1313
 FT CONFLICT 1335 1341
 FT CONFLICT 1368 1368
 FT CONFLICT 1470 1470

FT CONFLICT 1650 1650 W -> C (IN REF. 1).
 FT CONFLICT 1690 1690 A -> S (IN REF. 1).
 FT CONFLICT 1714 1714 G -> V (IN REF. 1).
 FT CONFLICT 1869 1869 E -> A (IN REF. 5).
 FT CONFLICT 1879 1883 GTPEY -> AHSRV (IN REF. 5).
 SQ SEQUENCE 2805 AA; 306392 MW; F3ED0A316599B2E CRC64;
 Alignment Scores:
 Pred. No.: 0.761 Length: 2805
 Score: 121.00 Matches: 77
 Percent Similarity: 33.86% Conservative: 31
 Best Local Similarity: 24.14% Mismatches: 133
 Query Match: 5.88% Indels: 78
 DB: 1 Gaps: 12

US-10-644-659A-1 (1-1146) x MAPA_HUMAN (1-2805)

QY 48 CCTCCGGAAGATCGCAC-----AGC 68
 Db 1781 ProProGluGluGluAspLysLeuThrArgSerProPheGluLeuLeuSerProProAla 1800
 QY 69 CACCTGTGTATCAGCTTGCCTGCGGAGTTG---GCAGCAGTGGCGGAATGAGAACAGCAT 125
 Db 1801 SerProGluMetValGlyGlnArgValProSerAlaProGlyGln-GluSerProI 1820
 QY 126 C-----AGGCAGGCCAGGACCTACA-----GGCTGGCTG-- 156
 Db 1820 eProAspProLysLeuMetProHisMetLysAsnGluProThrThrProSerTrpLeuAl 1840
 QY 157 -----CCGGAGGAGCCAGGACTCAGCTCAGCTCCTAAACCAATCACACCC 206
 Db 1840 aAspLeProProTrpValProLysAspArgProLeuProProAlaProGlySerPro-- 1859
 QY 207 TACTTCACACCAAGAGCTCAGAGTGCTCCCAAGATCGCCACCCCGCTGCCAGAGGACA 266
 Db 1860 -----AlaProGlyProProThrProAlaProGluSerHi 1871
 QY 267 TGGAGATGGACAAAGCTCAGAGAAAGCCCTGAGGTTCATCATCAAAAGAAAGAGGT 326
 Db 1871 sThrProAlaProPheSerTrpGlyThrProGluTyrAsp----- 1884
 QY 327 GTCCAAACCGTGTGTGACCAAGACTTACGAG-----AGAGGAGGGAGCGT 371
 Db 1885 -----SerValValAlaAlaValGlnGluGlyAlaAlaGluLeuGluGlyProTy 1902
 QY 372 GAGCCACTCAGCCAGAGTACGAGAGGGATGCTGTGTGCTTGAACCTGGGCGCAGCA 431
 Db 1902 rSerProLeuGlyLysAspTyrArgLysVal**GlyGluArgGluGluGlyArgAl 1922
 QY 432 GAATGACATTGACAAATCTCTCCACAGCCAGCGCTCCCAACCGGAGGAGAAATGTGC 491
 Db 1922 aGluAlaProAspLysSerSerHis***SerLysValProGluAlaArgLysSerHisAl 1942
 QY 492 CAACCTGGTGTCTGAGCTACCAAGGCTGGAGGCTGGAGAGGAGGAGGAGGAGGAGG 551
 Db 1942 aThrThrGluProGluGlnThrGlu-----ProGluGlnArgGluProThrPr 1958
 QY 552 GAGGAGTGCACCGGTAGACACAGAGAGCAGCGCTATGGAGGAGAGGCTGAGGAGGCGC 611
 Db 1958 oTyrProAspGluArgSerPheGln-TyrAlaAspIleTyrGluGlnMet----- 1974
 QY 612 CGAGCAGGATGGAGTGCAGGTGCTGTGTGCTCAGATCAAGCGCCCTTCCTCCAGGT 671
 Db 1975 -----MetLeuThrGlyLeuGlyProAlaCysProThrArgG 1987
 QY 672 AAACAGATTTACAGAGAAACTCAACTGCCAAAGCCCAAGAAATATAGCCAGTGGCGCA 731
 Db 1987 LuProProLeuGlyAlaAlaGlyAspTrpProCysLeuSerThrLysGluAlaAla 2007
 QY 732 CTTGAAGGAGATGGCAGCAGTGGCTGTGTGATGAACATACATCCAGAGCTCATCC 791
 Db 2007 laGlyArgAsnThrSerAlaGlu-----LysGluLeuSerSerProI 2021

Db 347 AlaGlnLeuGluMetAlaAlaGluValAlaAlaGlu-----ProMet 360
QY 961 GCTGCCACAGCAGGATGGCAAGATCCAGGTACTTTGAGATCTTTGACAGATAC 1020
Db 361 AlaArgAlaAaGAspAspAlaMetAsnAlaHisLeuLysGluVal----- 376
QY 1021 GTTCGTATTTCAGATAAAGTAGTGCGCATTCCTCGCTGCCAGGAACATGGAGTGTGA 1080
Db 377 LeuHisAlaAlaAaPProMetAlaAsnMetIleGlnLysArgArgAspThrAlaIle 396
QY 1081 GACTTTGAGGAGAGATGCTATGCGAAGCCGAGATGACCATGTTGTG 1128
Db 397 Asp---ArgGlyGluLeuValTyrProSerTyrHisGlyHisPheVal 411

RESULT 8
MICA_HUMAN
ID MICA_HUMAN STANDARD; PRT; 1067 AA.
AC Q8TDZ2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE CalL interacting molecule (NEDD9 interacting protein with calponin
DE homology and LIM domains).
GN NICAL OR MICAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A., SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND
INTERACTION WITH CASL AND VIMENTIN.
RX MEDLINE=21964028; PubMed=11827972;
RA Suzuki T., Nakamoto T., Ogawa S., Matsumura T., Tachibana K.,
RA Morimoto C., Hirai H.;
RT "MICAL, a novel CasL interacting molecule, associates with vimentin.";
RL J. Biol. Chem. 277:14933-14941(2002).
CC -!- FUNCTION: May be a cytoskeletal regulator that connects CasL/HEFL
CC to intermediate filaments.
CC -!- SUBUNIT: Associates with the SH3 domain of CasL/HEFL. Interacts
CC with vimentin.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: Expressed in the thymus, lung, spleen, kidney,
CC testis and hematopoietic cells.
CC -!- SIMILARITY: Contains 1 calponin-homology (CH) domain.
CC -!- SIMILARITY: Contains 1 LIM zinc-binding domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB048948; BAB86289.1; -;
CC Genew; HGNC:20619; NICAL.
CC MW; 607129; -;
DR InterPro: IPR001715; Calponin-like.
DR InterPro: IPR001781; LIM.
DR Pfam; PF00307; CH; 1.
DR Pfam; PF00412; LIM; 1.
DR ProDom; PD000094; LIM; 1.
DR SMART; SM00033; CH; 1.
DR SMART; SM00132; LIM; 1.
DR SMART; PS00021; CH; 1.
DR PROSITE; PS00478; LIM DOMAIN_1; 1.
DR PROSITE; PS00023; LIM DOMAIN_2; 1.
DR Cytoskeleton; LIM domain; Metal-binding; Zinc.
KW DOMAIN 508 609 CH.
FT DOMAIN 695 757 LIM.
FT DOMAIN 765 834 PRO-RICH.
SQ SEQUENCE 1067 AA; 117932 MW; 8A65FD13E72B60E6 CRC64;

Alignment Scores: 1.02 Length: 1067
Pred. No.: 118.50 Matches: 101
Score: 32.04% Conservative: 56
Percent Similarity: 20.61% Mismatches: 136
Best Local Similarity: 5.76% Indels: 197
Query Match: 1 Gaps: 25
DB:

US-10-644-659A-1 (1-1146) x MICA_HUMAN (1-1067)

QY 31 GGCCCCAGCCCAAGAGCGCCCTCCGGAAGATACGCACAGCCACCCCTGTCTATCAGC---TTG 87
Db 619 GlyProValSerGlnAlaSerProGlyThrSerSerAlaValLeuPheLeuSerLysLeu 638
QY 88 GCCCGAGGTGGCAGCAG---TGGCGAATGAGAACAGC----- 123
Db 639 GlnArgThrLeuGlnArgSerArgAlaLysGluAsnAlaGluAspAlaGlyLysLys 658
QY 124 ATCAGG-----CAGGCCAGAGCCTACAGGTGGCTGCCGGAGGAGCCAGGAC 174
Db 659 LeuArgLeuGluMetGluAlaGluThrProSerThrGluValPro-----Pro 674
QY 175 TCACCTCAAGCTCTAAACCAATCACACCCCTACTTCACACGAGAAGCTCAGAGT--- 231
Db 675 AspProGluProGlyValProLeuThrProProSerGlnHisGlnGluAlaGlyAlaGly 694
QY 231 ----- 231
Db 695 AspLeuCysAlaLeuCysGlyGluHisLeuTyrValLeuGluArgLeuCysValAsnGly 714
QY 232 -----GCCCAAGTCGCCACCC 249
Db 715 HisPhePheHisArgSerCysPheArgCysHisThrCysGluAlaThrLeuTyrProGly 734
QY 250 CGCCTGCCAGAGCAGATGAGATGGACAA-----AGCTCAGAGAAGGCCCTGAGGTT 303
Db 735 GlyTyrGluGlnHisProGlyAspGlyHisPheTyrCysLeuGlnHisLeuProGlnThr 754
QY 304 TCTCACATCAAAAAGAGAGGTGTCCAAAACGGTGTCCAGCAAGACTTACGAGAGGGA 363
Db 755 AspHis-----LysLysGluGlySer-----AspArgGly 764
QY 364 GGGNACGTGAGCCACCTCAGCCACAGTACGAGAGGAT----- 402
Db 765 ProGluSerProGluLeuProThrProSerGluAsnSerMetProProGlyLeuSerThr 784
QY 403 -----GCTGCT---GTGCTTGAACCTGGCGACCCAGAGAAATAC 438
Db 785 ProThrAlaSerGlnGluGlyAlaGlyProValProAspProSerGlnProThrArgArg 804
QY 439 ATTGACAGAACTCTCCACAGCCACGGCTCCCAACGGCGGAGG----- 480
Db 805 GlnIleArgLeu-----SerSerProGluArgGlnArgLeuSerSerLeuAsn 820
QY 481 -----AGAAAATGTGCCAACCTG 498
Db 821 LeuThrProAspProGluMetLeuProProLysProProArgSerCysSerAlaLeu 840
QY 499 GTGCTCTGAGCTAACCAAG-----GGCTCGAGAGTGTATGCAGCAGGAGGAGCC 546
Db 841 AlaArgHisAlaLeuGluSerSerPheValGlyTyrGlyLeuProValGlnSerProGln 860
QY 547 ACATGGAGGAGTGACAGCGTAGACACAGAGCAGCGGTATGGAGGAGAGGCTGAGGAG 606
Db 861 AlaLeuValAlaMetGluLysGluGluLysGluSerProPheSerSerGluGluGlu 880
QY 607 AGG-----CCCAGCAGGATGAGTGCAGGTGGGTGTG----- 639
Db 881 GluAspValProLeuAspSerAspValGluGlnAlaLeuGlnThrPheAlaLysThrSer 900
QY 640 -----GTCAGGATCAAGCGCCCTTGCCTCC 666

```
Db 901 GlyThrMetAsnAsnTyProThrTrpArgThrLeuLeuArgAlaLysGlu 920
Qy 667 CAGGTAAACAGATTACAGAGAACTCAACTGCAAGCCCAA----- 708
Db 921 GluMetLysArgPhe-----CysLysAlaGlnThrIleGlnArgLeu 935
Qy 708 ----- 708
Db 936 AsnGluIleGluAlaLeuArgGluLeuGluAlaGluGlyValLysLeuGluLeuAla 955
Qy 709 -----CAGAAATATATCCAGTGGCGCAACTTGAAA-----GGGAGATGGCAG 750
Db 956 LeuArgGlnSerSerSerProGluGlnGlnLysLysLeuTrpValGlyGlnLeuLeu 975
Qy 751 CAGTGGGCTGATGAAC-----ATACAATCC 777
Db 976 GlnLeuValAspLysLysAsnSerLeuValAlaGluAlaGluLeuMetIleThrVal 995
Qy 778 CAGAGCTCATCTCTTCAGTGAAGAGTTGATTACGAGCTGCCATGTCACCCGCTTA 837
Db 996 GlnGluLeuAsn-----LeuGluGluLysGlnTrpGlnLeuAspGlnGluLeuArg--- 1012
Qy 838 CACAAAGAGATGAGGCTATGGCGCCCAAGAGAAACCAAACTGCTGAAGGGCC 897
Db 1013 -----GlyTyrMetAsnArgGluGluAlaLeuLysThrAla---AlaAsp 1026
Qy 898 AAGCTCTGAGGAGCAGATCTACAGGAATGATGACATGTGCTTCATTATCTGCACA 957
Db 1027 ArgGlnAlaGluAspGlnValLeuArgLysLeuValAspLeuVal----- 1041
Qy 958 ATGGCTGCCACAGCAGATGCGCAAGATCCAGCTTACTTTTGGAGATCTCTTTGACAGA 1017
Db 1042 -----AsnGlnArgAspAlaLeuLysLeuArg-----PheGlnGlu 1052
Qy 1018 TACCTTGCTATTTCAGATTAAGTAGTGGGC 1047
Db 1053 GluArgArgLeuSerGluLeuAlaLeuGly 1062

RESULT 9
ID IF2_SNNY3 STANDARD; PRT; 1001 AA.
AC P75889;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Translation initiation factor IF-2.
GN INFB OR SLR0744.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugitani M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res 3:109-136(1996).
CC -!- FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex [by similarity].
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the IF-2 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90900; BAA16696.1; --
DR PIR; S74544; S74544.
DR HAMAP; MF_00100; 1.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR000178; IF2.
DR InterPro; IPR006847; IF2_N.
DR InterPro; IPR005225; Small_GTP.
DR InterPro; IPR009000; Translat_factor.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 2.
DR Pfam; PF04760; IF2_N; 2.
DR PRINTS; PR00315; ELONGATNFCT.
DR ProDom; PD186100; IF2; 1.
DR TIGRFAMs; TIGR00487; IF-2; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS01176; IF2; 1.
DR Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 496 648 G-DOMAIN.
FT NP_BIND 509 509 GTP (BY SIMILARITY).
FT NP_BIND 552 556 GTP (BY SIMILARITY).
FT NP_BIND 606 609 GTP (BY SIMILARITY).
SQ SEQUENCE 1001 AA; 108118 MW; 4C7F00DE56CE31C CRC64;
```

```
Alignment Scores:
Pred. No.: 1.39 Length: 1001
Score: 116.50 Matches: 81
Percent Similarity: 33.61% Conservative: 39
Best Local Similarity: 22.69% Mismatches: 111
Query Match: 5.66% Indels: 126
DB: 1 Gaps: 18
```

US-10-644-659a-1 (1-1146) x IF2_SNNY3 (1-1001)

```
Qy 12 CGAAAAGAAAGAGCGGGAGGGCCGCAAGAGCC-----CTCCGGAAGATACG 62
Db 89 LysGlnGluLysSerGlyGlyProSerProAlaArgProThrProProArgProLys 108
Qy 63 CACAGCCACCCCTGGTCATCAGTTGCCCGAGG---TTGGCAGCAGTGGCGCAATGAGAA 119
Db 109 LeuGlnAlaProLysAlaProThrProProGlnProProValAlaLys-AlaSerAlaPr 128
Qy 120 CAGCATCAGGCAGGCCCGCAGGAGCTACAGGCTGGCTGCCGGAGGGACCCAGGACTCAC 179
Db 128 clyslleGlnLysGlnGluPro-----AlaG1 138
Qy 180 TCAGCTCTTAACCAATCACACCCCTACT-----TC 212
Db 138 nGluAlaProLysSerValAlaProProThrGlnProLeuAlaProProValProSe 158
Qy 213 ACACCAAGAAAGCTCAGAGTGGCCCAAGTCGCA----- 246
Db 158 rLeuGlnSerProProSerLysProAlaProProThrProProAlaLysAlaAlaPr 178
Qy 247 -----CCCCGCTCCAGAGGACATGGAGTGAAGAAAGCTCAGAAAGCCCTCAGGT 302
Db 178 cAlaProArgLeuAlaGlyProProGlyArgThrAlaSerProAsnLysThrAlaValPr 198
Qy 303 TTCTCATCAATAAAGAGAGGTGCCAAACGGTGGTCAGCAAGACTTACGAGAGAGG 362
Db 198 cAlaProAlaLysProLysValAsnArgProGluIleValSerLeuLysAspAsnArgG1 218
Qy 363 A-----GGGACCTGAGCCACCTCAGCCACGAGTACGAGAGGATGCT-- 405
Db 218 yGlnAlaArgSerProGlyAsp-----ArgGluGluLysValAla1 232
Qy 406 -GGTGTCTTGAACCTGGGCGAGGAGATGATGACATTCAGAGAAATCTCCACAGCCACGG 464
```

```

Db 232 eAlaalaProGluProProLysProLysValGluLeuArgarg-----Pr 247
QY 465 CTCCCAACGCGGAGGAGAAATGTGCCAAGCTGTGTCTGAGCTAAACCAAGGGCTGGAG 524
Db 247 oLysProProArgProGluGluAspGluAsnLeuProGluLeuLeu----- 262
QY 525 AGTGATGGAGCAGGAGGAGCCACATGGAGGAGTACACAGCGTAGACACAGAGGACGCGG 584
Db 263 -----GluPheProLeuSerArgGlyValAspGlyAspAsnAspAl 279
QY 585 CTATGAGGAGAG-----GCTGAGAGAGGCGCCGAGCAGGAGTGAGTGCGGTGGTGT 638
Db 279 aAspAspGlyAspLeuLeuSerThrGluLysProLysPro----- 292
QY 639 GGTGAGGATCAGGCGCCCTGCTCCCTCCAGGTAAACAGATTACAGAGAACTCAACTG 698
Db 293 -----LysLeuLysAspGProThrProProArgLeu----- 302
QY 699 CAAAGCCCAACAGAAATATAGCCAGTGGCGCACTTGAAGGAGATGCGACATGGCG 758
Db 303 -----GlyLysProAspGlnTrpG1 309
QY 759 TGATGAACATACATCCAGAGCTCAAT----- 789
Db 309 uAspAsp-----GluAspGluLysAlaAsnLysAlaLysAlaAsnLysGlyLysAr 327
QY 790 -----CCTTTTCAGTGAAGAGTTGATTACGAGTGGCGCTGCCCGCTACACAAAGG 845
Db 327 gArgProLysMetAspAspAspAspAspLeuAspLeuAsp-----G1 342
QY 846 AGATGAGGGC-----TATGGCGCGCC----- 867
Db 342 yAspAsnGlyProLysProThrLeuValSerLeuSerIleAlaArgProProLysProLy 362
QY 868 -----AAAGAGGAGAACCAAACTGCTGAAGGCGCAAGCGT 903
Db 362 sSerLeuAlaAlaLysProSerThrProThrValAlaLysValLysLys 378

RESULT 10
NCR2 MOUSE
ID NCR2 MOUSE STANDARD; PRT; 2472 AA.
AC Q9WU42; Q9WU43; Q9WUC1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Nuclear receptor co-repressor 2 (N-CoR2) (Silencing mediator of
DE retinoic acid and thyroid hormone receptor) (SMRT) (Thyroid-,
DE retinoic-acid-receptor-associated co-repressor) (TR3 receptor-
DE associating factor) (TRAC).
GN NCoR2 OR SMRT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS ALPHA AND BETA).
RC TISSUE=Brain, and Spleen;
RA MEDLINE=99178941; PubMed=10077563;
RX Ordentlich P., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.;
RT "Unique forms of human and mouse nuclear receptor corepressor SMRT.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644 (1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
RC TISSUE=Embryo;
RX MEDLINE=99192215; PubMed=10097068;
RA Park E.J., Schreen D.J., Yang M., Li H., Li L., Chen J.D.;
RT "SMRte, a silencing mediator for retinoid and thyroid hormone
RT receptors-extended isoform that is more related to the nuclear
RT receptor corepressor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524 (1999).
RN [3]
RP INTERACTION WITH HDAC7.

```

```

RX MEDLINE=20107033; PubMed=10640276;
RA Kao H.-Y., Downes M., Ordentlich P., Evans R.M.;
RT "Isolation of a novel histone deacetylase reveals that class I and
RT class II deacetylases promote SMRT-mediated repression.";
RL Genes Dev. 14:55-66 (2000).
CC -1- FUNCTION: Mediates the transcriptional repression activity of some
CC nuclear receptors by promoting chromatin condensation, thus
CC preventing access of the basal transcription.
CC -1- SUBUNIT: Forms a large corepressor complex that contains SIN3A/B
CC and histone deacetylases HDAC1 and HDAC2. This complex associates
CC with the thyroid (TR) and the retinoid acid receptors (RAR) in the
CC absence of ligand and may stabilize their interaction with TFIIB.
CC Interacts with HDAC10 and MINT (By similarity). Interacts with
CC HDAC7.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Alpha;
CC IsoId=Q9WU42-1; Sequence=Displayed;
CC Name=Beta;
CC IsoId=Q9WU42-2; Sequence=VSP_003414;
CC -1- TISSUE SPECIFICITY: Ubiquitous. Also widely expressed in early
CC embryos.
CC -1- DOMAIN: The N-terminal region contains repression functions that
CC are divided into three independent repression domains (RD1, RD2
CC and RD3). The C-terminal region contains the nuclear receptor-
CC interacting domains that are divided in two separate interaction
CC domains (ID1 and ID2).
CC -1- DOMAIN: The two interaction domains (ID) contain a conserved
CC sequence referred to as the CORNR box. This motif is required and
CC sufficient to permit binding to unliganded TR and RARs. Sequences
CC flanking the CORNR box determine nuclear hormone receptor
CC specificity.
CC -1- SIMILARITY: Contains 1 SANT-A domain.
CC -1- SIMILARITY: Contains 1 MYB-like domain.
CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/ebis/
CC or send an email to license@ebi.ac.uk).
CC -----
EMBL; AF113001; AAD20944.1; -
EMBL; AF113002; AAD20945.1; -
EMBL; AF125671; AAD22972.1; -
TRANSFAC; T04690; -
DR MGD; MGI:1337080; Ncor2.
DR InterPro; IPR001005; MYB DNA binding.
DR Pfam; PF00249; myb DNA-binding; 2.
DR SMART; SM00717; SANT; 2.
DR PROSITE; PS50090; MYB_3; 1.
KW Nuclear protein; transcription regulation; DNA-binding; Repressor;
KW Coiled coil; Alternative splicing.
FT DOMAIN 165 207 COILED COIL (POTENTIAL).
FT DOMAIN 254 312 INTERACTION WITH SIN3A/B (BY SIMILARITY).
FT DNA_BIND 429 474 SANT-A (POTENTIAL).
FT DNA_BIND 609 653 MYB.
FT DOMAIN 492 560 COILED COIL (POTENTIAL).
FT DOMAIN 652 682 COILED COIL (POTENTIAL).
FT DOMAIN 775 804 PRO-RICH.
FT DOMAIN 989 999 PRO-RICH.
FT DOMAIN 1351 1357 PRO-RICH.
FT DOMAIN 2094 2098 CORNR BOX OF ID1.
FT DOMAIN 2296 2300 CORNR BOX OF ID2.
FT DOMAIN 494 507 POLY-GLN.
FT DOMAIN 1615 1619 POLY-ALA.
FT DOMAIN 2434 2437 POLY-PRO.
FT VARSPPLIC 36 254 Missing (in isoform Beta).
FT /FTID=VSP_003414.

```


RA Grandy D.K., Leduc R., Makam H., Flanagan T., Diliberto E.J.,
 RA Thomas G., Civelli O., Viveros O.H.,
 RL Submitted (Oct-1990) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 634-646.
 RC TISSUE: Adrenal chromaffin;
 RX MEDLINE=95262699; PubMed=7744058;
 RA Strub J.-M., Garcia-Sablone P., Lonning K., Taupenot L., Hubert P.,
 RA van Dorsselaer A., Anis D., Metz-Boutigue M.-H.;
 RT "Processing of chromogranin B in bovine adrenal medulla.
 RT Identification of secretoryin, the endogenous C-terminal fragment of
 RT residues 614-626 with antibacterial activity.";
 RL Eur. J. Biochem. 229:356-368(1995).
 RN [5]
 RP CHARACTERIZATION OF SECRETOLYTIN.
 RX MEDLINE=96184581; PubMed=8603705;
 RA Strub J.-M., Hubert P., Nulians G., Anis D., Metz-Boutigue M.-H.;
 RT "Antibacterial activity of secretolytin, a chromogranin B-derived
 RT peptide (614-626), is correlated with peptide structure.";
 RL FEBS Lett. 379:273-278(1996).
 CC -!- FUNCTION: Secretogranin I is a neuroendocrine secretory granule
 CC protein, which may be the precursor for other biologically active
 CC peptides. The 16 pairs of basic AA distributed throughout its
 CC sequence may be used as proteolytic cleavage sites.
 CC -!- FUNCTION: Secretolytin has antibacterial activity.
 CC -!- SUBCELLULAR LOCATION: Neuroendocrine and endocrine secretory
 CC granules.
 CC -!- PTM: O-glycosylated (Probable).
 CC -!- SIMILARITY: Belongs to the chromogranin / secretogranin protein
 CC family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X55027; CAA38846.1; -;
 DR EMBL; U88551; AAC48720.1; -;
 DR EMBL; X55489; CAA39109.1; -;
 DR PIR; S15901; S15901.
 DR InterPro; IPR001819; Chromogranin_AB.
 DR InterPro; IPR001990; Granin.
 DR Pfam; PF01271; Granin; 1.
 DR PRINTS; PR00659; CHROMOGRANIN.
 DR PROSITE; PS00422; GRANINS 1; 1.
 DR PROSITE; PS00423; GRANINS 2; 1.
 DR Sulfation; Cleavage on pair of basic residues; signal.
 KW SIGNAL
 FT CHAIN 1 20
 FT PEPTIDE 21 646 SECRETOGGRANIN I.
 FT PEPTIDE 418 484 GAWK PEPTIDE.
 FT PEPTIDE 634 646 SECRETOLYTIN.
 FT DISULFID 36 57 BY SIMILARITY.
 FT MOD_RES 158 158 SULFATION (POTENTIAL).
 FT MOD_RES 315 315 SULFATION (BY SIMILARITY).
 FT CONFLICT 64 64 N -> S (IN REF. 1).
 FT CONFLICT 70 70 N -> D (IN REF. 2).
 FT CONFLICT 93 98 SEAPGL -> FRSPRAS (IN REF. 3).
 FT CONFLICT 181 181 T -> M (IN REF. 2).
 FT CONFLICT 261 261 H -> R (IN REF. 2).
 FT CONFLICT 386 386 P -> R (IN REF. 2).
 FT CONFLICT 481 481 H -> L (IN REF. 3).
 FT CONFLICT 597 597 M -> V (IN REF. 2).
 SQ SEQUENCE 646 AA; 73339 MW; 420DB1178FD9E415 CRC64;
 Alignment Scores:
 Pred. No.: 1.68 Length: 646
 Score: 115.00 Matches: 81
 Percent Similarity: 34.99% Conservative: 53
 Best Local Similarity: 21.15% Mismatches: 121
 Query Match: 5.59% Indels: 128

DB: 1 Gaps: 21
 US-10-644-659A-1 (1-1146) x SGI_BOVIN (1-646)
 QY 13 GAAAGGAAAGCGGGAGGCGCCAGCAGAGC----- 45
 Db 169 GluGluArgLeuSerGluGlyProGlyLysAlaGlnThrAlaPheLeuAsnGlnArgAsn 188
 QY 46 -----GCCCTCCGGAAGATACGACACAGCCACCCCTGGTTCATCAGCTGGCCCGAGGT 96
 Db 189 GlnThrProAlaLysLysGluGluValSerArgTyrAspThrGlnSerAlaArgGly 208
 QY 97 TGGCAGCAG-----TGGGCGAATGAGAACAGCAGCAGCCAGGAGCCCTACA----- 147
 Db 209 LeuGluLysSerHisSerArgGluArgSerSerGlnGluSerGlyGluThrLysSer 228
 QY 148 -----GGCTGG----- 165
 Db 229 GlnGluAsnTrpProGlnGluLeuGlnArgHisProGluGlyGlnGluAlaProGlyGlu 248
 QY 166 ACCCAGGAC-----TCACCTCAAGCTCTAAACCAATCACACCCCTACTTCCACACCAG 219
 Db 249 SerGluGluAspAlaSerProGluValAspLysArgHisSerArgPro---ArgHisHis 267
 QY 220 AAAGCTCAGACTGCCCAAG-----TCGCCACCCCGCTGCCAGAA 261
 Db 268 HisGlyArgSerArgProAspArgSerSerGlnGluArgLysAsnProLeuGluGlu 287
 QY 262 GGACAT---GGAGATGGACAAAGC---TCAGAGAAAGCCCTGAGGTTTCTCATCAATCAA 315
 Db 288 SerHisValGlyThrGlyAsnSerAspGluGluLysAlaArgHisProAlaHisPheArg 307
 QY 316 AAGAAAGAGGTGTCAAACCGTGTGTCAGCAAGCTTACGAGAGAGAGGGAGCTGAGC 375
 Db 308 AlaLeuGlu-----GluGlyAlaGluTyrGlyGluGluVal--- 319
 QY 376 CACCTCAGCCACAGTACGAGAGGATGCTGGTGTGTTGAACCTGGGCGAG----- 426
 Db 320 -----ArgArgHisSerAlaAlaGlnAlaProGlyAspLeuGlnGly 333
 QY 427 -----CCAGAGAAT 435
 Db 334 AlaArgPheGlyGlyArgGlyGlyGluHisGlnAlaLeuArgArgProSerGluGlu 353
 QY 436 GACATTGACAGATCTCCACAGCCAGGCG---TCCCCACCGCGGAGGAGAAATGTGCC 492
 Db 354 SerLeuGluGlnGluAsnLysArgHisGlyLeuSerProAsp----- 367
 QY 493 AACCTGGTGTCTGAGCTAACCAAGGCTGAGAGAGTGTGATGGAGCAGGAGGAG- 543
 Db 368 -----LeuAsnMetAlaGlnGlyTyrSerGluGluSerGluGluArgGlyPro 384
 QY 544 -----CCACATGGAGG-----AGTGACAGCGTA 567
 Db 385 AlaProGlyProSerTyrArgAlaArgGlyGlyGluAlaAlaTyrSerThrLeuGly 404
 QY 568 GACACAGAGCAGCGGCTATGTGAGGAGAGGCTGAGGAGAGCGCCGAGCAGGATGGATG 627
 Db 405 GlnThrAspGluLysArgPheLeuGlyGluThrHisHisArgValGlnGluSerGlnArg 424
 QY 628 CAGGTGGTGTGGTTCAGGATCAAGCGCCCTTGGCTCCCGAGTAACAGATT----- 681
 Db 425 AspLysAla-----ArgArgArgLeuProGlyGluLeuArgAsnTyrLeuAsp 440
 QY 682 -----ACAGAGAACTCACTGCAAGCCCAACAGAAATATAGCCAGTGGGCACTTG 735
 Db 441 TyrGlyGluGluLysGlyGluGluAlaAlaArgGlyLysTrpGlnProGlnGlyAspPro 460
 QY 736 AAAGGG-----AGATGCGACAGTGGCGGTGATAA 765
 Db 461 ArgAspAlaAspGluAsnArgGluAlaArgLeuArgGlyLysGlnTyrAlaProHis 480
 QY 766 CACATACATCCAGAG-----CTCAATCTCTTC----- 795

Db 481 HisileThrGluYsArgLeuGlyGluLeuLeuAsnProPheYrAspProSerGlnTrp 500
QY 796 ---AGTGAAGAGTTTATTACGAGCTGGCATGTCCACCGCTACACAAAGAGATGAG 852
Db 501 LysSerSerArgPheGluArgLysAspProMetAspSerPheLeuGluGlyGluGlu 520
QY 853 GGCTATGGC 861
Db 521 GLuAsnGly 523

RESULT 12

TRP_DROME STANDARD; PRT; 1275 AA.
AC P19334;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transient receptor potential protein.
GN TRP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCB1_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=50180449; PubMed=2516726;
RA Montell C., Rubin G.M.;
RT "Molecular characterization of the Drosophila trp locus: a putative
RL integral membrane protein required for phototransduction.";
RN Neuron 2:1313-1323(1989).
RP [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=50148782; PubMed=2482778;
RA Wong F., Schaefer E.L., Roop B.C., Lamendola J.N., Johnson-Seaton D.,
RT "Proper function of the Drosophila trp gene product during pupal
RT development is important for normal visual transduction in the
RT adult.";
RL Neuron 3:81-94(1989).
RN [3]
RP SEQUENCE OF 1126-1275 FROM N.A.
RX MEDLINE=89042982; PubMed=3118483;
RA Wong F., Yuh Z.T., Schaefer E.L., Roop B.C., Ally A.H.;
RT "Overlapping transcription units in the transient receptor potential
RT locus of Drosophila melanogaster.";
RL Sonat. Cell Mol. Genet. 13:661-669(1987).
CC -!- FUNCTION: REQUIRED FOR PHOTOTRANSDUCTION. SUGGESTED TO MEDIATE
CC CALCIUM ENTRY. SEEMS TO FORM A LIGHT-SENSITIVE CALCIUM PERMEANT
CC CHANNEL.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE RHABDOMERIC
CC MEMBRANES OF THE PHOTORECEPTOR CELLS.
CC -!- SIMILARITY: Belongs to the transient receptor family. StrpC
CC subfamily.
CC -!- SIMILARITY: Contains 2 ANK repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M34394; AAA28976.1; -;
CC EMBL; M21306; AAA56928.1; -;
CC EMBL; M18634; AAA28977.1; -;
CC FlyBase; FBgn0003861; trp.
CC GO; GO:0016028; C:ribosome; IDA.
CC GO; GO:0015279; F:store-operated calcium channel activity; NAS.

DR GO; GO:0008377; P:light-induced release of calcium, from inte. .; IDA.
DR GO; GO:0009416; P:response to light; IMP.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR002111; Cat_channel_TrpL.
DR InterPro; IPR005821; Ion_trans.
DR InterPro; IPR002153; Trans_receptor.
DR InterPro; IPR004729; Trp_CaChannel.
DR Pfam; PF00023; ank; 2.
DR Pfam; PF00520; Ion_trans; 1.
DR PRINTS; PRO1097; TRNSRECEPTR.
DR SMART; SMO0248; ANK; 2.
DR TIGRFAMs; TIGR00870; trp; 1.
DR PROSITE; PS50297; ANK_REPEAT; 1.
DR PROSITE; PS50297; ANK_REPEAT; 1.
KW Ionic channel; Transmembrane; Ion transport; Calcium channel;
KW Vision; ANK repeat; Repeat.
FT DOMAIN 1 366 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 367 387 POTENTIAL.
FT DOMAIN 388 390 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 391 411 POTENTIAL.
FT DOMAIN 412 418 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 419 439 POTENTIAL.
FT DOMAIN 440 450 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 451 471 POTENTIAL.
FT DOMAIN 472 507 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 508 528 POTENTIAL.
FT DOMAIN 529 541 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 542 562 POTENTIAL.
FT DOMAIN 563 638 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 639 659 POTENTIAL.
FT DOMAIN 660 1275 EXTRACELLULAR (POTENTIAL).
FT REPEAT 69 98 ANK 1.
FT REPEAT 143 172 ANK 2.
FT CONFLICT 285 288 QGRQ -> ASSE (IN REF. 2).
FT CONFLICT 326 329 RRQK -> POE (IN REF. 2).
FT CONFLICT 365 374 KPQVKFTHS -> NPLSSSRTP (IN REF. 2).
FT CONFLICT 755 785 S -> N (IN REF. 2).
SQ SEQUENCE 1275 AA; 142589 MW; 91CFCD9896989B1 CRC64;

Alignment Scores:

	Pred. No.:	Score:	Length:
	2.11	114.00	1275
	Percent Similarity:	36.52%	Matches: 86
	Best Local Similarity:	24.16%	Conservative: 44
	Query Match:	5.54%	Mismatches: 116
	DB:	1	Indels: 110
			Gaps: 21

US-10-644-659a-1 (1-1146) x TRP_DROME (1-1275)

QY	9	GGGCGAAAAGGAAGCGGGGCGCCAGC-----	38
Db	994	GlyLysProGlyThrMetGlyLysProThrAspLysLysProGlyAspLysAsp 1013	
QY	39	CAAGAGCGCCCTCCGGAAGATACGACAGCCACCCCTGGTCATCAGCTGGCCCGAGTTG 98	
Db	1014	LysGlnGlnProLysAspSerLysProSerAlaGly-----GlyProLys--- 1029	
QY	99	GCAGCAGTGGGCGAATGAGAACAGCATCAGGCAGCGCCGAGCGCTACAGGCTGCTGCC 158	
Db	1030	-----ProGlyAspGlnLysProThr-Pr 1037	
QY	159	GGGAGGAGCCAGGACTCACCTCAAGCTCTTAACCAATCACACCCCTACTTCACACA 218	
Db	1037	OGlyAlaGlyAlaProLysProGlnAlaAlaGlyThrIleSerLysProGlyGluSerGl 1057	
QY	219	GAAGCTCAGAGTGGCCCAAGTCGCCACCCCGCTGCCAGAGGACATGGAGAT----- 273	
Db	1057	nLys---LysAspAlaPro---AlaProProThrLysPro-----GlyAspThrLys 1072	
QY	274	-----GGCAAAAGCTCAGCAAAAGCCCTGAGGTTTCTCATCAAA 314	
Db	1072	sProAlaAlaProLysProGlyGluSerAlaLys-----ProGluAlaAla---AlaLys 1089	

```
QY 315 AAAGAAAGAGGTGTCACAAACGGTGTGTCAGCAACAGTTCACGAGAGGAGCGGACGCTGAG 374
Db |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||
1099 slysgluuserSerlyThrGluAseRlyProAlaAaThraSncGlyAlaAlay 1109
QY 375 CCACCTCAGCCACAGGTACGAGAGGATGCT-----GGTGTGCTTGAACCTGGCA 425
Db |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||
1109 sSerAlaAlaProSerAlaProSerAlaAlaLysProAspSerlySLeuLysProGly-- 1128
QY 426 GCCAGAGAATGACATTGACAGAACTCCTCCACGACCGGCTCCCA-----ACGGGGAG 479
Db |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||
1129 -----AlaAlaGlyAlaProGluAlaThrLysAl 1138
QY 480 GAGAAATGTCACCACTGCTGTGTCAGTCAACCAAGGGCTGAGAGTGTGAGCAGGAGCA 539
Db |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||
1138 aThrAsnGlyAlaSerLysProAspGlyLysSerGly-----ProGl 1153
QY 540 GGAGGCC-----ACATGAGGAGTCAACAGCTAGACACAGAGGACAG 581
Db |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||
1153 uGluProLysLysAlaAlaGlyAspSerLysProGlyAspAlaLysAspLysAspLys 1173
QY 582 CGGTATCGAGGAGGCTGAGGAGAGCCCGCAGCAGGATGGAGTGGCTGTGGT 641
Db |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||
1173 slySProGlyAspAspLysAspLysLysProGlyAspAsp----- 1186
QY 642 CAGGATCAAGCGCCCTTCCCTCCAGGTAAACAGATTACAGAGAACTCAACTCA 701
Db |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||
1187 -LysAspLysLysProAlaAspAsnAspLysLys-----Pr 1199
QY 702 AGCCCAACAGAAATATAGCCAGTGGCGCACTTAAAGGAGGATGGCAGGCTGCTGA 761
Db |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||
1199 alaAspAspLysAspLysLysProGlyAspAspLys-----As 1212
QY 762 TGAACACATCAATCCCAAGAGCTCAATCTTTCAGTCAAGAGTTTGTATCAGGCTGGC 821
Db |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||
1212 pLySProGlyAspAspLysAspLysLysProSerAspLysAsp-LysLys---- 1230
QY 822 CATGTCCACCGCTTACACAGAGAGATGAGGGGTATGGCGGCCCAAGAGACCA 881
Db |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||
1231 -----ProAlaAspAspLysAspLysLysProAlaAlaProLeuLys--ProA 1247
QY 882 AACTGTGAAAGCGGCGAGCTGCTG-----AGGAGCAGCATCTA 920
Db |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||
1247 laileLysValGlyGlnSerAlaAlaAlaGlyGlyGluArgGlyLysSerThrValT 1267
QY 921 CAGGGAATGATGACATGCTCTTCATTCGTGCAATGGCTC 964
Db |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||
1267 hrglyArg-----MetileSerGlyTrpLeu 1275

RESULT 13
TCOF HUMAN
ID TCOF HUMAN STANDARD; PRT; 1411 AA.
AC Q13428; Q99408; Q99860;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Treacle protein (Treacher Collins syndrome protein).
GN TCOF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96154183; PubMed=8563749;
RA Dixon J., Gladwin A.J., Dixon M.J., Loftus S.K.,
RA Bonner C.A., Koprivnikar K., Wasmuth J.J.;
RT "Positional cloning of a gene involved in the pathogenesis of
RT Treacher Collins syndrome";
RL Nat. Genet. 12:130-136 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97228900; PubMed=9074926;
RA Dixon J., Edwards S.J., Anderson I., Brass A., Scambler P.J.,
```

RA Dixon M.J.;
RT "Identification of the complete coding sequence and genomic
organization of the Treacher Collins syndrome gene";
RL Genome Res. 7:223-234 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97250498; PubMed=9096354;
RA Wise C.A., Chiang L.C., Paznekas W.A., Sharma M., Musy M.M.,
RA Ashley J.A., Lovett M., Jabs E.W.;
RT "TCOF1 gene encodes a putative nucleolar phosphoprotein that exhibits
mutations in Treacher Collins syndrome throughout its coding
region";
RT Proc. Natl. Acad. Sci. U.S.A. 94:3110-3115 (1997).
RN [4]
RP VARIANTS LEU-439; VAL-810; VAL-1313 AND GLY-1355, AND VARIANT TCS
ARG-53.
RX MEDLINE=97195537; PubMed=9042910;
RA Edwards S.J., Gladwin A.J., Dixon M.J.;
RT "The mutational spectrum in Treacher Collins syndrome reveals a
predominance of mutations that create a premature-termination
codon";
RT Am. J. Hum. Genet. 60:515-524 (1997).
RL Am. J. Hum. Genet. 60:515-524 (1997).
CC -!- DISEASE: Defects in TCOF1 are the cause of Treacher Collins
syndrome (TCS) [MIM:154500]. TCS is an autosomal dominant disorder
of craniofacial development that occurs with an incidence of
1/50,000 live births. The clinical features of TCS are bilaterally
symmetrical and include: (1) abnormalities of the external ears,
atresia of the external ear canals, and malformation of the middle
ear ossicles, which may result in conductive hearing loss; (2)
lateral downward sloping of palpebral fissures, frequently with
colobomas of the lower eyelids; (3) hypoplasia of the mandible and
zygomatic complex; (4) cleft palate.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U40847; AAC50903.1; -
DR EMBL; U76366; AAC51181.1; -
DR EMBL; U84664; AAC51185.1; JOINED.
DR EMBL; U84640; AAC51185.1; JOINED.
DR EMBL; U84641; AAC51185.1; JOINED.
DR EMBL; U84642; AAC51185.1; JOINED.
DR EMBL; U84643; AAC51185.1; JOINED.
DR EMBL; U84644; AAC51185.1; JOINED.
DR EMBL; U84645; AAC51185.1; JOINED.
DR EMBL; U84646; AAC51185.1; JOINED.
DR EMBL; U84647; AAC51185.1; JOINED.
DR EMBL; U84648; AAC51185.1; JOINED.
DR EMBL; U84649; AAC51185.1; JOINED.
DR EMBL; U84650; AAC51185.1; JOINED.
DR EMBL; U84651; AAC51185.1; JOINED.
DR EMBL; U84652; AAC51185.1; JOINED.
DR EMBL; U84653; AAC51185.1; JOINED.
DR EMBL; U84654; AAC51185.1; JOINED.
DR EMBL; U84655; AAC51185.1; JOINED.
DR EMBL; U84656; AAC51185.1; JOINED.
DR EMBL; U84657; AAC51185.1; JOINED.
DR EMBL; U84658; AAC51185.1; JOINED.
DR EMBL; U84659; AAC51185.1; JOINED.
DR EMBL; U84660; AAC51185.1; JOINED.
DR EMBL; U84661; AAC51185.1; JOINED.
DR EMBL; U84662; AAC51185.1; JOINED.
DR EMBL; U84663; AAC51185.1; JOINED.
DR EMBL; U79659; AAB40722.1; -
DR EMBL; U79645; AAB40722.1; JOINED.
DR EMBL; U79646; AAB40722.1; JOINED.
DR EMBL; U79647; AAB40722.1; JOINED.

```

1150 SerSerProSerValSerSerThrLeuAlaAlaLysAspAspProAspGlyLysGlnGlu 1160
274 GGACAAAGCTCAGAGAAAGCCCTGAGGTCTTCATCATCAAAAAGAAAGAGGTGTCCAAA 333
1170 AlaLysProGlnGlnAlaAlaGlyMetLeuSer-----ProLysThrGlyGlyLys 1185
334 ACGGTGGTCAGCAAGACTTTACGAGAGAGAGGGGACGTGAGCCACTCAGCCACACAGGTAC 393
1187 GluAlaAlaSer-----GlyThrThrProGlnLysSerArgLysPro 1200
394 GAGAGGGATGCTGGT----- 408
1201 LysLysGlyAlaGlyAsnProGlnAlaSerThrLeuAlaLeuGlnSerAsnIleThrGln 1220
409 -----GTGCTTGAACCTGGCGACCCAGACAGAAATGACATTCACAGA 447
1221 CysLeuLeuGlyGlnProTriProLeuAsnGlnAlaGlnValGlnAlaSerValValLys 1240
448 ATCCCTCCACACGCCACCGCTCCCAACGGCGAGGAGAAAAATGTGCCAACCTGGTGTCTGAG 507
1241 ValLeuThrGluLeuLeuGlnGlnArgLysLysValValAspThrThrLysGluSer 1260
508 CTAACAAGGGCTGG-----AGAGTGATGGACGACGAGGAGAGCCC----- 546
1261 SerArgLysGlyTriPgluSerArgLysArgLysLeuSerGlyAspGlnProAlaAlaArg 1280
547 ACATGGAGAGTGACAGCGCTAGACACAGAGGACGCGGTATGGAGGAGAGGCTGAGGAG 606
1281 ThrProArgSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1300
607 AGGCCCGACGACGATGGAGTGAGTGCAGGTGGCTGTGTCAGGATCAAGCGCCCTTGCCTCC 666
1301 SerProGluLysThrSerThr----- 1307
667 CAGGTAAACAGATTTTACAGAGAACTCAACTGCCAAAGCCCAACAGAAATATAGCCAGTG 726
1308 -----ThrSerLysGlyLysAlaLysArgAspLysAlaSerGlyAspVal 1322
727 GGCAACTGTAAGAGGAGATGGCAGCAGTGGGTGATGAACAATCAATCAATCCAGAAAGCTC 786
1323 LysGluLysLysGlyLysGlySerLeuGlySer-----GlnGlyAlaLysAsp 1338
787 AATCCTTTTCAGTGAAGAGTTTCATTACGAGCTGGCCATGTCACCGCCCTACACAAGGA 846
1339 GluPro-----GluGlnGlnGlnLysGlyMetGlyThr---ValGluGlyGly 1354
847 GATGAGGGGTATGGCGGCCCCAAGAGGAACAAACTGTGTAAGGGCCCAAGCCTGCT 906
1355 AspGlnSerAsnProLysSerLysLysGluLysLysLysSerAspLysArgLysLysAsp 1374
907 GAGGAG 912
1375 LysGlu 1376

RESULT 14
XPC_DROME STANDARD; PRT; 1293 AA.
ID XPC DROME
AC Q24595;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-repair protein complementing XP-C cells homolog (Xeroderma
DE pigmentosum group c complementing protein homolog) (XPCCM) (Mutagen-
DE sensitive 209 protein).
DE MUS210 OR XPC OR XPC.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DF CN BW; TISSUE=Embryo;

```

RX MEDLINE=94173659; PubMed=8127661;
RA Henning K.A., Peterson C., Leigerski R., Friedberg E.C.;
RT "Cloning the Drosophila homolog of the Xeroderma pigmentosum
RT complementation group C gene reveals homology between the predicted
RT human and Drosophila polypeptides and that encoded by the yeast RAD4
RT gene.";
RL Nucleic Acids Res. 22:257-261(1994).
CC -1- FUNCTION: Involved in DNA excision repair. May play a part in DNA
CC damage recognition and/or in altering chromatin structure to
CC allow access by damage-processing enzymes (By similarity).
CC -1- FUNCTION: Involved in nucleotide excision repair of DNA damaged
CC with UV light, bulky adducts, or cross-linking agents.
CC -1- SUBUNIT: Heterodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: Belongs to the XPC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z28622; CAA82262.1; -;
DR PIR; S42402; S42402.
DR FlyBase; FBgn0004698; mus210.
DR InterPro; IPR004583; Rad4.
DR Pfam; PF03835; Rad4; 1.
DR TIGRfam; TIGR00605; rad4; 1.
KW DNA repair; DNA-binding; Nuclear protein.
FT DOMAIN 33 40 POLY-ASP.
FT DOMAIN 632 637 POLY-SER.
FT DOMAIN 686 693 POLY-SER.
FT DOMAIN 922 938 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 1195 1211 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 1275 1291 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ SEQUENCE 1293 AA; 144177 MW; 99DF671F9A4151G5 CRC64;
Alignment Scores:
Pred. No.: 2.28 Length: 1293
Score: 113.50 Matches: 83
Percent Similarity: 39.18% Conservative: 60
Best Local Similarity: 22.74% Mismatches: 157
Query Match: 5.52% Indels: 65
DB: 1 Gaps: 16
US-10-644-659A-1 (1-1146) x XPC_DROME (1-1293)
QY 7 CCGGGCGAAGGAAGCGG-----CAGGGCCCGCCAGAGCGCC 48
Db 551 PrAlaGluLysGluSerSerLysSerThrIleSerLysGluAlaGluLysLysAsnAsn 570
QY 49 CTCGGGAAGATACGACAGCCACCCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGAGTGG 108
Db 571 AlAlaLysLysAlaGluAlaLysProLeuSerLysSerThrThrLysGly----- 586
QY 109 GCGAATGAGACAGCATCAGGAGGCGGAGGACCTACAGGTGGCTGGCGGAGGAGAC 168
Db 587 -----SerGluThrThrLysSerGlyThrValProLysValLys 599
QY 169 CAGGACTCACCTCAAGCTCCTCAACCAATCACACCCCTACTTCACACAGAAAGCTCAG 228
Db 600 LysGluLeuSerLeuSerSerLys---LeuValGluLysSerLysHisGlnLysAlaHis 618
QY 229 AGTGCCCAAGTCGCCACCCCGCTGCCGAGAGGACATGGAGATGGACAAAGCTCA--- 285
Db 619 ThrSerSerLysSerAspThrSerPheAspGluLysProSerThrSerSerSerLys 638
QY 286 -----GAGAAGCCCTGAGGTTCTCATCATCAAAAGAAAGAGGTGTCCAAACG 336
Db 639 CysLeuLysGluGluTyrSerGluLeuGly---LeuSerLysLysLeuLeuLysProThr 557

QY 337 GTGCTCAGCAAGACTTACGAGAGAGAGGAGGAGCGTACGCCACCTCACCACAGGTACGAG 396
Db 658 LeuSerSerLysLeuValLeuLysSerLysAsnGlnSerPheSerSerAsn---Lys 676
QY 397 AGGATGCTGTGTGCTTGAACCTGGCAGCAGAGAAATGACATTGACAGATCCTCCAC 456
Db 677 SerAspThrSerPheGluGlu-----AsnProSerThr----- 687
QY 457 AGCCACGGCTCCCAACGGGAGGAGAAATGTCCAACTGGTGTCTGAGTAAACCAAG 516
Db 698 SerSerSerLysSerLeuLysGluThrAlaLysLeuSerSerLysLeuGlu 707
QY 517 GCCTGGAGAGTATGGAGCAGGAGGCCACATGGAGGAGTGACAGC----- 564
Db 708 AspLysLysValAlaSerSerAlaGluThrLysThrLysValGlnSerSerLeuLys 727
QY 565 ---GTAGACACAGAGGACGCGGTATGGAGAGAGCTGAGGAGGCGCCGAGCAGAT 621
Db 728 ArgValThrThrGlnAsnIleSerGluSerGlyAspSerLysLysSer----- 743
QY 622 GGATGTCAGTGGCTGTGGTCAAGATCAAGCGCCCTTGCCTCCAGGTAAACAGATTT 681
Db 744 -----LysValAlaProValAspThrPheSerProValAlaGlyArgThrArgAla 761
QY 682 ACAGAGAACTCAACTGCAAAAGCCCAAGAAATATAGCCAGTGGGCAAC----- 732
Db 762 ThrValLys-----ProLysThrGluLysProGlnValValGlySerProValIle 779
QY 733 -----TTGAAAGGAGATGCGCAGCTGGGCTGATGACACATACATCCCAAG 783
Db 780 ProLysLeuMetLeuSerLysValLysGlnLeuAsnAlaLysHisSerAspThrGluAsn 799
QY 784 CTCATCTCTTCAGTGAAGAGTGTGATTACGAGCTGGCCATGTCCACCCGCTACACAA 843
Db 800 AlaSerProAlaAsnLysHisLeuGlnGluGln-----Arg-AsnThrAr 814
QY 844 GGATGAGGCTATGGCCGCCCAAGAGAGAAACCAAACTGCTGAAAGGCGCAAGCGT 903
Db 814 gGluThrArgSerArgSer-----LysSerProLysValLeuLysSerProSerPh 831
QY 904 GCTGAGGAGCATCTACAGGAAATGATGAGACATGTCTTATTATCGACAATGGCT 963
Db 831 eLeuLysLysSerSerAspGlyAlaAspSerThrSerAlaProGlnLysHisGlnMetA 851
QY 964 CGCCACAGACAGATGGCAA---GATCCAGGTACTTTTGGAGATCTCTT-----TGAC 1014
Db 851 laProGluThrLysAlaArgIleSerProAsnPheLeuSerGluAlaLeuProAlaArg 871
QY 1015 AGATACGTTCC 1025
Db 871 InLeuArgSer 874
RESULT 15
VSX1 CHICK
ID VSX1 CHICK STANDARD; PRT; 350 AA.
AC Q9IAL2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Visual system homeobox 1 (transcription factor VSX1) (Homeobox protein
DE Chx10-1).
GN VSX1 OR CHX10-1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20108593; PubMed=10640715;
RA Chen C.-M.A. Cepko C.L.;
RT "Expression of Chx10 and Chx10-1 in the developing chicken retina.";
RL Mech. Dev. 90:293-297(2000).

Search completed: May 4, 2004, 08:52:39
Job time : 38 secs